

BIRLA CENTRAL LIBRARY

PILANI [ RAJASTHAN ]

Class No. 620.1128

Book No. F892I

Accession No. 76309





**The Inelastic Behavior  
of Engineering Materials  
and Structures**



# The Inelastic Behavior of Engineering Materials and Structures

**Alfred M. Freudenthal**

PROFESSOR OF CIVIL ENGINEERING  
COLUMBIA UNIVERSITY  
LECTURER AND RESEARCH CONSULTANT  
IN THEORETICAL AND APPLIED MECHANICS  
UNIVERSITY OF ILLINOIS

**John Wiley & Sons, Inc., New York**  
**Chapman & Hall, Limited, London**

Copyright, 1950

by

John Wiley & Sons, Inc.

---

*All Rights Reserved*

*This book or any part thereof must not  
be reproduced in any form without the  
written permission of the publisher.*

*To the memory of*  
**MY FATHER**



## PREFACE

With the boundaries of physical science expanding rapidly, present-day engineering students must necessarily assimilate information and knowledge of subjects that continue to become more and more complex.

In the study of the nature and mechanical behavior of engineering materials, however, their knowledge seldom progresses beyond the level of the elementary theory of elasticity and strength. This approach to instruction, where all materials are considered to be homogeneous, isotropic, and linearly elastic continuous media, is most expedient, since, without it, practical stress analysis and the design of structural members and machine parts would be impossible. However, the implications and consequences of the approximations so introduced are easily forgotten, and as a result, these extremely useful tools of definitely limited applicability are frequently accepted as a true reflection of the properties of all engineering materials. Furthermore, this acceptance by students in the early college years shows considerable vitality in later years even when confronted with engineering reality.

Perhaps it is not surprising then that the approach to many solutions of problems of strength and structural design does not transcend the concept of the resistance of the continuous elastic medium. Too often in engineering practice an attempt is made to interpret the mechanical behavior of real materials solely in terms of elastic theory rather than to consider the nature of the material and its actual response to forces as the manifestation of its heterogeneous internal structure. One example of such a misdirected attempt is found in the proposed extension of the concept of the elastic stress concentration around a hole to cracks of submicroscopic size, in order to explain the difference between the theoretical (atomic) and the technical cohesive strength of materials. Moreover the discrepancy between the elaborate analysis in the aerodynamical phase of the design of modern airplanes, or the thermodynamical analysis in the design of gas

turbines, and the crudeness of the concepts and methods used in proportioning the parts of such machines, emphasizes the need for a more effective approach to the study of the behavior of real materials.

Most of our knowledge concerning the mechanical behavior of engineering materials is derived from phenomenological (large-scale) observations and, frequently, an empirical interpretation of a large volume of partially coordinated observed facts. The shortcomings of this approach become more and more evident as the materials increase in complexity, as the conditions of service under which the materials will perform become more diversified, and, finally, as the number of "laws" that must be devised to express the interrelation between the relevant variables necessarily increases. In general, such "laws" are valid only within the range of the observations that serve as a basis for their derivation.

Hence, the volume of experimental work required to provide the empirical "laws" needed to solve the increasingly complex problems of engineering design is becoming especially unwieldy. So much so, in fact, that experimental research is in danger of defeating its own purpose because the newly acquired factual knowledge can be neither coordinated with existing knowledge nor effectively distributed and utilized. Any worker in this field of engineering would be hard pressed indeed to keep abreast of what is being done in his special field, and still less in any neighboring field, in spite of its possible importance to his own work. Moreover, both research personnel and research facilities are heavily overtaxed.

The most effective remedy for this condition would be the development of a more fundamental approach in the research of the behavior of engineering materials; an approach based on the analysis of the underlying unifying principles by which the relevant engineering concepts could be clarified, broadened, and developed; an approach with the purpose of providing the guiding principles for further experimental research which would advance organized knowledge rather than merely accumulate a rather uncoordinated mass of facts. It is this necessity for *economy of thought* leading automatically to *economy of experiment* that is expressed by the old saying that "nothing is more practical than a good theory."

The unifying principles, by which the apparently complex phenomenological behavior of real materials can be interpreted in terms of a few basic concepts, are the laws governing the formation of matter from particles and larger structural elements at different levels of aggregation. Thus the deformational properties of single crystals are closely related to the principles governing the formation of the particular type of crystal out of atoms, ions, or molecules. Likewise, the response to applied load of polycrystalline metal aggregates or of high polymers is essentially determined by the laws of formation of such materials from single crystals or giant molecules.

The development of new engineering material, as well as the variation in service requirements of modern engineering machinery and structures, has created many problems which can only be handled by considering the internal structure of the materials and its response to applied forces and conditions. For example, the assumptions of homogeneity, isotropy, and time-independent elasticity are irreconcilable with the phenomena of fatigue, time- and temperature-sensitive cohesive strength, and creep of real materials. An analysis of the mechanical behavior of such materials requires a consideration of their structure and their phenomenological response interpreted in terms of structural changes under the applied forces and conditions.

Thus, the large-scale behavior becomes predictable, at least qualitatively, from a knowledge of the internal structure. The laws governing the aggregation of matter, which are involved in the interpretation of its structure and of the resulting mechanical properties are few and relatively simple. They are concerned mainly with the inherent mobility of the constituent particles or structural elements and with the forces of interaction between them at the different levels of aggregation. The adoption by the engineer of structural considerations to supplement and amplify the purely phenomenological approach, which must remain the principal engineering procedure, will considerably simplify and advance the development of phenomenological (engineering) methods, by providing the necessary guidance for the establishment of real functional relations of rather broad validity between the relevant variables, rather than the empirical relations of strictly limited validity obtained in most engineering tests.

In this book an attempt has been made to present an integrated

picture of the response to applied forces of engineering materials, through a combination of the structural (physical) and the phenomenological (engineering) approaches. A principal difficulty in preparing what is intended primarily as a textbook for graduate students but also should be useful for research workers and designers has been the huge volume of relevant material and its diversity. Moreover, the variety of existing factual information, as well as the variety of opinions and theories, could not be permitted to obscure the fundamental principles on which a unified presentation of the subject matter could be based. Therefore, if the purpose of the book was to be achieved to any appreciable extent, it was necessary to impose restrictions on both the breadth and depth of the treatment of the various phenomena and aspects.

It is for the same reason that a detailed description and discussion of the experimental techniques and specific conclusions of various experimenters have been omitted. As an introductory textbook rather than a reference book, the main emphasis has been placed on the presentation of the physical response of engineering materials to forces, time, and temperature. Every effort has been made to present the information in a form comprehensible to a reader with an engineering background. This point of view has also influenced the selection of the literature cited in the text. References have been so selected as to enable the reader to find a few of the most important or most representative original publications concerning any particular topic. This sampling of the literature should provide sufficient further references if more intensive study is desired.

An introductory chapter discusses the reason for studying mechanical behavior and examines the basic concepts and definitions. Parts A and B develop the structural and phenomenological framework of the theory of inelasticity, which is the theory of general deformational behavior of engineering materials. Part C deals with selected problems of the mechanics of the inelastic continuum, with the design of engineering structures, and with mechanical testing.

This book is based essentially on mimeographed notes prepared for a two-semester course for graduate students at the University of Illinois and later repeated at Columbia University. It could not have been written without the generous and enthusiastic

support of Professor F. B. Seely, Head of the Department of Theoretical and Applied Mechanics, University of Illinois. His encouragement and criticism were a continuous source of inspiration during the work on the notes and on the manuscript. The author is deeply indebted to Professor Seely as well as to the members of the department with whom many of the chapters were discussed and who offered valuable suggestions and criticism. He also wishes to express his appreciation to the office staff of the department for their cooperation.

ALFRED M. FREUDENTHAL

*May, 1950*



# CONTENTS

## INTRODUCTION

1. Purpose of the Study of Mechanical Properties of Materials . . . . .	1
2. Basic Concepts and Definitions . . . . .	6
The material body . . . . .	6
Forces . . . . .	8
Deformation . . . . .	9
3. Methods of Approach . . . . .	11
Levels of approach . . . . .	11
Rheological theory . . . . .	14
Mechanical behavior of groups of elements . . . . .	15
Energy considerations . . . . .	20
4. Elastic and Inelastic Behavior . . . . .	21

## PART A

### THE STRUCTURAL ASPECT OF MECHANICAL BEHAVIOR

#### CHAPTER 1 GENERAL CONCEPTS OF THE STRUCTURE OF MATTER

5. Oscillators and Quanta . . . . .	29
Oscillators: concept of atomic structure . . . . .	29
Two-oscillator model . . . . .	30
Quanta . . . . .	32
6. Bohr's Model of the Atom . . . . .	35
The "solar-system" model . . . . .	35
Periodic table of elements. Pauli's exclusion principle . . . . .	37
7. Quantum Statistics and "Wave Particles" . . . . .	43
The wave function . . . . .	43
The statistical concept . . . . .	46
8. The Size of Atoms and Molecules . . . . .	52

#### CHAPTER 2 THE STRUCTURE OF MATTER

9. Forces of Interaction . . . . .	61
Types of bonds . . . . .	61
Ionic bonds . . . . .	63
Covalent bonds . . . . .	64
Metallic bonds . . . . .	68
Complex bonds . . . . .	71
Intermolecular bonds . . . . .	73
Interatomic and intermolecular forces . . . . .	73
10. Thermal Oscillations . . . . .	76
11. Ordered and Unordered State . . . . .	84

12. Structural Geometry of the Ordered State . . . . .	88
13. Finite Groups. Formation of Real Materials . . . . .	94
14. Microscopic and Macroscopic Structure . . . . .	105
<b>CHAPTER 3 STRUCTURAL THEORIES OF DEFORMATION</b>	
15. Statistical Aspect of the Behavior of an Assembly of Particles . . . . .	109
16. "Spontaneous" Changes of State . . . . .	116
17. Forced Change of State (Inelastic Deformation) . . . . .	122
18. Geometrical Aspect of Plastic Deformation . . . . .	128
19. Structural Theory of Inelastic Deformation . . . . .	135
20. Complex Types of Deformation . . . . .	140
Texture formation. Change in molecular structure . . . . .	140
Strain hardening . . . . .	142
Creep . . . . .	148
Classification of mechanisms of change of state . . . . .	151
21. Volumetric Deformation . . . . .	153
22. Fracture . . . . .	156
<b>PART B</b>	
<b>MECHANICS OF INELASTIC DEFORMATION</b>	
<b>CHAPTER 4 MECHANICAL VARIABLES</b>	
23. Mechanical State . . . . .	167
24. Dynamical Variables . . . . .	170
25. Kinematical Variables . . . . .	177
26. Mechanical Variables in Curvilinear Coordinates . . . . .	185
<b>CHAPTER 5 THE MECHANICAL EQUATION OF STATE</b>	
27. Change of State. Thermodynamic Considerations . . . . .	188
28. Change of State. Energy of the Deformable Body . . . . .	198
29. Superposition of Simplified Equations of State . . . . .	200
30. Linear Equations . . . . .	204
31. Application of the Concept of the Equation of State . . . . .	208
32. Limit of Continuous Change of State. Fracture . . . . .	211
<b>CHAPTER 6 LINEAR BEHAVIOR</b>	
33. Elasticity. Small Strain . . . . .	214
34. High Elasticity. Finite Strain . . . . .	217
35. Linear Bodies. Systematics of Inelastic Behavior . . . . .	220
Integration of basic equations . . . . .	220
Elastic-viscoelastic analogy . . . . .	229
36. Mechanical Models. Simple Behavior . . . . .	232
37. Mechanical Models. Complex Behavior . . . . .	235
38. Superposition Theories . . . . .	243
39. Nonlinear Mechanical Models . . . . .	246
<b>CHAPTER 7 PLASTICITY</b>	
40. Initiation of Plastic Deformation . . . . .	249
41. Conditions of Plasticity . . . . .	254

42. Theories of Plastic Deformation . . . . .	263
43. Residual Stresses following Plastic Deformation . . . . .	273

**CHAPTER 8 WORK HARDENING OF POLYCRYSTALLINE METALS**

44. Structural Theories of Work Hardening . . . . .	281
45. A General Law of Work Hardening . . . . .	287
46. Time Effects. Thermal Stability of Work Hardening . . . . .	297
47. Phenomenological Analysis of Plastic Deformation with Work Hardening . . . . .	301

**CHAPTER 9 CREEP AND RELAXATION**

48. Creep of Viscoelastic Materials . . . . .	305
49. Creep of Metals . . . . .	312
50. Analysis of Creep Problems . . . . .	319
51. Relaxation . . . . .	322

**CHAPTER 10 INELASTIC BEHAVIOR UNDER DYNAMIC CONDITIONS**

52. The Damping Capacity . . . . .	326
53. Theory of Anelastic Effects . . . . .	339
54. Impact . . . . .	349

**CHAPTER 11 FRACTURE**

55. Theories of Brittle Fracture . . . . .	357
56. Ductile and Brittle Fracture . . . . .	368
57. Transition Temperature . . . . .	374
58. Fracture under Single-Stroke, Sustained, or Repeated Load . . . . .	380
59. Condition of Fracture . . . . .	387
60. Fracture on Release of Load . . . . .	394

**CHAPTER 12 RHEOLOGICAL BEHAVIOR OF SUSPENSIONS AND GELS**

61. Brownian Motion. Thixotropy . . . . .	398
62. Viscosity and Yield Limit . . . . .	402
63. Rheological Measurement. Behavior of Technical Materials . . . . .	408

**PART C****APPLICATION OF THE MECHANICS OF INELASTIC BEHAVIOR****CHAPTER 13 PLASTICITY. PROBLEMS OF EQUILIBRIUM**

64. Torsion . . . . .	417
65. Bending. Nonuniform Stress . . . . .	420
66. The Thick-Walled Cylinder under External and Internal Pressure . . . . .	430
67. The Rotating Cylinder and Disk . . . . .	440
68. The Blunted Wedge . . . . .	444
69. Contact Pressure . . . . .	448

**CHAPTER 14 PLASTICITY. PROBLEMS OF FLOW**

70. Two-Dimensional Problems. Glide Lines . . . . .	452
71. Technological Problems. Pressing and Rolling . . . . .	455

Compression of a thin sheet . . . . .	455
Rolling of a thin strip . . . . .	457
<b>72. Technological Problems. Wire Drawing . . . . .</b>	<b>460</b>
<b>CHAPTER 15 WORK HARDENING AND CREEP. SPECIAL PROBLEMS</b>	
<b>73. Bending . . . . .</b>	<b>464</b>
Work hardening . . . . .	464
Creep . . . . .	466
<b>74. Infinitely Long Cylindrical Hole under Pressure within Infinite Plane . . . . .</b>	<b>468</b>
Work hardening . . . . .	468
Creep . . . . .	469
<b>75. Wire Drawing . . . . .</b>	<b>470</b>
<b>CHAPTER 16 DESIGN FOR PLASTICITY AND WORK HARDENING</b>	
<b>76. Limitations of Design for Elasticity . . . . .</b>	<b>472</b>
<b>77. The Factor of Safety . . . . .</b>	<b>477</b>
<b>78. Design for Plasticity of Redundant Engineering Structures (Theory of Limit Design) . . . . .</b>	<b>481</b>
<b>79. Procedures of Limit Design . . . . .</b>	<b>489</b>
Single load application . . . . .	490
Repeated application of the same load . . . . .	491
Alternating application of different loads or mobile loads .	494
Analysis of redundant elastic-plastic system . . . . .	495
Consideration of the actual form of the stress-strain diagram . . . . .	496
<b>80. Theory of Limit Design and Experiment . . . . .</b>	<b>496</b>
<b>81. Stability in Compression within the Elastic-Plastic Range . . . . .</b>	<b>500</b>
<b>CHAPTER 17 DESIGN FOR CREEP</b>	
<b>82. Metallic Structures and Parts at High Temperatures . . . . .</b>	<b>506</b>
<b>83. Viscoelastic Materials . . . . .</b>	<b>515</b>
Columns . . . . .	518
Arches . . . . .	519
<b>84. Concrete and Reinforced Concrete . . . . .</b>	<b>523</b>
<b>CHAPTER 18 SIGNIFICANCE OF MECHANICAL TESTING. INTERPRETATION OF RESULTS</b>	
<b>85. Mechanical Tests and Inelastic Behavior. Machine Effects . . . . .</b>	<b>527</b>
<b>86. The Hardness Test . . . . .</b>	<b>534</b>
<b>87. The Tension Test . . . . .</b>	<b>540</b>
<b>88. The Torsion Test . . . . .</b>	<b>554</b>
<b>89. Impact Tests . . . . .</b>	<b>555</b>
<b>90. Fatigue Tests . . . . .</b>	<b>560</b>
<b>AUTHOR INDEX . . . . .</b>	<b>575</b>
<b>SUBJECT INDEX . . . . .</b>	<b>579</b>

## INTRODUCTION

### **1. Purpose of the Study of Mechanical Properties of Materials**

The formulation of the relations between the forces acting on a material body under various conditions and the resulting deformational response of the material constitutes the subject of the study of mechanical properties of materials. It is the main purpose of this book to treat this subject primarily with respect to the deformational response designated as *inelastic*.

Technological progress depends on the development of new materials and on the more effective use of existing ones. No matter how rapid scientific progress is nor how significant new developments in science are, their technological impact depends on materials of required properties being available to build the necessary equipment.

The most important group of properties are the mechanical properties. Although the properties essential to the purpose and the performance of a certain material or apparatus are frequently nonmechanical, the conditions of existence of such apparatus or material presuppose the presence of definite mechanical properties. Thus in the manufacture of materials for whatever use—metals for load-carrying structures, armor plate, electric wire, or watchsprings; high polymers for tires, insulators, or photoelastic studies; or highly deformable substances such as paints and glues—the creation of certain mechanical properties must be one of the principal objectives. No other properties, no matter how essential they may be to the purpose of the material, can become significant unless the mechanical properties that are required to resist the acting forces have been brought into existence.

Frequently certain observed mechanical properties are not of direct importance, in controlling either the necessary resistance to fracture or the deformability under service loads. Neverthe-

less, their relation to the desired nonmechanical properties may provide a relatively simple indirect indication of those properties. In other cases, the influence of the atomic or molecular structure of a material on certain mechanical properties is known from experiment; it may then be possible to derive information concerning the internal structure of the material from observation of its mechanical properties. For example, viscosity measurements of colloidal suspensions are used to obtain information concerning the form and concentration of the suspended particles;<sup>1-1</sup> such inference is based on theoretically established relations between these structural characteristics and the coefficient of viscosity of the suspension. On the other hand, measurements of the viscosity of lubricants have been found useful in classifying their performance in service, although no clear functional relation between lubricating effect and viscosity has so far been established.

The study of mechanical properties of materials may thus have one of three purposes:

(a) To investigate mechanical properties that are or are assumed to be directly needed for the contemplated use of the material;

(b) By observing certain mechanical properties, to derive information about other properties, mechanical or nonmechanical, which cannot easily be observed directly, the assumption being that a correlation, functional or empirical, can be established between the observed and the inferred properties;

(c) To correlate observed mechanical properties with theoretical concepts concerning those properties, which are derived from the consideration of the atomic, molecular, or microscopic structure of the material; the aim of such study is to establish methods by which the character of the internal structure may be inferred from observations of mechanical properties and, vice versa, mechanical properties predicted from (or "interpreted" in terms of) the internal structure.

Of the three purposes specified, the greatest difficulties are encountered in investigations of problems arising under purpose *a*, although, on first consideration, this group of problems appears to be the most simple. However, the selection of measurable mechanical properties which are relevant with regard to the contemplated use of the material, and their correlation with the

actual performance of the material in service are generally not easy problems. It is usually found easier to produce materials with certain specified properties than to specify the properties or the measurable characteristics required for a given purpose.

One of the principal difficulties encountered in the study of the mechanical properties of materials consists in making observations of the properties under conditions in which the material is actually being used. This difficulty is created by the fact that, during the process of making the "observation" of a property in a mechanical test of the material, the "observed" property or characteristic is usually changed. Thus the observed and measured properties are those of the material changed by the test, not of the material in its initial condition or in its condition of service. Only the elastic properties are unaffected by the test; it is therefore within the elastic range only that the performance of materials in tests and under service conditions can be easily correlated.

Problems that arise under purpose *b* include the majority of engineering problems. They are relatively simple if a definite relation exists between the investigated property and the actually observed property as, for example, between the "ultimate tensile strength" observed in the conventional tensile test and the "hardness" observed in the indentation test. The required relation can be reliably established only if both properties can be directly observed and measured. In this case the problem is created by the fact that it is easier (or less expensive) to measure the related property than the property that is considered significant with regard to the contemplated use of the material. If both properties are not directly observable, the existence of the necessary relation between them can be deduced only from circumstantial evidence or by logical inference; this relation is therefore considerably less reliable.

Frequently the derived property is not even well defined, as in the investigation of the "ductility" or the "toughness" of a metal. In spite of their frequent use in engineering practice, these terms designate rather vague characteristics of structural performance. The attempted correlation with an observable simple characteristic of the tension test, such as the elongation at fracture, must therefore remain unsatisfactory; such correla-

tion does not make the fundamental concept more definite. The principal difficulty in problems arising under purpose *b* is therefore the establishment of the necessary relation.

The problems arising under purpose *c* are those concerning the interrelation of the internal structure of materials with their mechanical behavior. Whereas problems under *a* and *b* are essentially engineering problems, problems under *c* are on the border line between the fields of study of physics and physical chemistry on the one hand and that of engineering on the other. It is, however, on the satisfactory solution of this type of problems that engineering progress in the development of new materials largely depends. Unless the physical basis of the origin of mechanical properties is more thoroughly known, and this knowledge is correlated with the results of observations of the large-scale deformational response to loads, progress in the development of engineering materials and in their evaluation for various purposes will be extremely difficult. This statement becomes obvious if an interpretation of the results of such assumedly simple engineering tests as the creep-strength test or the fatigue test is attempted which would go beyond the simple establishment, by curve fitting, of the empirical relation between fracture stress and time or the number of load cycles sustained, valid for the particular conditions of the tests. Neither of those tests can be interpreted in terms that would make possible an extrapolation of the test results beyond these particular conditions. The general significance of the results is lost unless the interrelation between the observed large-scale behavior of the material and the internal structure in terms of atomic, molecular, or microscopic phenomena can be established.

An understanding of the structure of matter is therefore not only of theoretical, but also of immediate practical importance in all engineering problems concerned with materials, their manufacture, development, testing, and use. Familiarity with the modern physical concepts of atomic structure, of statistical mechanics, and of the basically discontinuous (quantum) nature of all processes of energy transformation is a necessity for the engineer who is attempting to improve his materials and really to understand their properties. Unless the engineer knows how the material is built up, he will hardly be able to understand what makes it deform and break. And as the chemists have learned

to understand that a chemical formula is not more than an empirical statement of observed facts unless it is interpretable in structural terms, that is, in terms of the organization of the minute particles of matter, so the engineer must realize that mechanical properties of materials, such as deformability and resistance to fracture under various conditions, require for their real understanding an interpretation in terms of the internal structure of the material. This does not mean, however, that the large-scale or "engineering" approach to the study of mechanical behavior is not very useful for certain purposes; it only means that this approach does not lead to such an interpretation of the observed behavior, which would make possible the prediction of the general behavior under various conditions. The relationships obtained in the engineering approach are empirical rather than functional; their validity is therefore limited to the specific testing conditions for which they have been established.

In studying mechanical properties the factor of time must be taken into account. Although it is generally realized that mechanical properties vary with temperature, their change with time is frequently disregarded. However, the concept of the *stability* of mechanical properties is one of great practical importance. This stability which is a function of the (thermal) stability of the internal structure is governed by the second law of thermodynamics. The longer a material has existed, the nearer its structure has approached the conditions of maximum stability towards which it tends after infinite time. Thus, the older a material, the more homogeneous its internal structure, and the more stable are its mechanical properties. The mechanical properties of natural stone are therefore not appreciably affected by the passage of a time period that would be sufficient to cause substantial changes in the mechanical properties of a metal and, still more, of a high polymer such as rubber. Every process by which the internal structure of the material is changed in the direction of either higher or lower thermal stability, such as heat-treatment or permanent deformation imposed in the course of fabrication processes, affects therefore not only the mechanical properties themselves but also their rate of change with time. Under certain conditions, such as occur in creep and fatigue, this rate of change may be of engineering significance.

## 2. Basic Concepts and Definitions

The mechanics of deformable bodies is concerned, in general, with the relations between the forces acting on a material body and the resulting deformational response (state of motion) in the body. In the mechanics of the ideally *elastic* material the deformational response is fully recoverable on release of the forces; the only material characteristic of the body is its elasticity. In the mechanics of the ideally *viscous* and the ideally *plastic* material the release of the forces is not accompanied by a deformational response; the deformation caused by the forces is fully irrecoverable. In the mechanics of deformable bodies of *general* behavior both of these deformational responses are present in varying proportion, depending on the nature of the material, on the manner in which the forces are applied (especially as influenced by time), and on temperature.

The relations between the acting forces and the resulting state of motion are derived from the observation and analysis of the deformational response of bodies of finite dimensions assumed to be homogeneous. This approach to the study of deformable bodies is called *phenomenological*, because the phenomena are treated as they appear in large-scale observation; it leads to phenomenological theories of mechanical behavior.

The three fundamental physical concepts involved in the study of the mechanics of deformable bodies are those of (*a*) the material body, (*b*) the forces acting on it, and (*c*) the resulting motion or deformation. All other concepts can be derived from these three concepts, which therefore are discussed in greater detail.

**THE MATERIAL BODY.** An aggregation of matter having a definite spatially limited shape is usually called a material body. Such a body is defined by its geometrical and its material properties; together they produce its mechanical properties. The body is limited by the surface, which determines its geometrical shape.

A material body consists of *structural elements*. Each element may be considered to be a smaller body which, in turn, is built up of still smaller bodies. These bodies of different order of magnitude represent the structural elements of consecutive orders of magnitude. For reasons of expediency this divisibility of

matter is arbitrarily limited at a certain level, in accordance with the purpose of the particular investigation.

Thus in a study of the deformational behavior of a polycrystalline metal body under the action of forces it will, in general, be sufficient to consider the crystals as the smallest "indivisible" elements, whereas in the study of changes of deformational behavior resulting from temperature or metallurgical conditions, such as precipitation, diffusion, or recrystallization, the atoms represent the "indivisible" elements. In the study of piezoelectric and dielectric properties even this division is insufficient; the particles that compose the atom (the nucleus and the electrons) have to be introduced as the "indivisible" elements of the body. Finally, if the effects of nuclear radiation on mechanical properties are investigated, the divisibility of the nucleus itself must be considered.

The mechanical properties of the material body are made up of the geometrical and the material properties of the structural elements at the different levels of their aggregation. The geometrical properties are defined by the geometrical shape of the elements and by their arrangement in space, whereas the material properties are related to both their mass and the so-called "mass defect." This is the mass "lost," that is, transformed into the energy that binds the elements together when structural elements at any level of aggregation are combined into units of the next higher order of magnitude; the mass of the composite body is smaller than the sum of the masses of the individual elements forming it. The resulting "mass defect" is equivalent to the binding (cohesive) energy of the body.<sup>2-1</sup> The helium atom, for instance, has a mass of 4.004 atomic weight units, whereas the mass of the particles forming it is  $4 \times 1.008 = 4.032$  atomic weight units. Thus the mass defect associated with the formation of helium out of fundamental particles is 0.028 atomic weight units. The energy transformed into binding energy is obtained by multiplying the mass defect by the square of the velocity  $c$  of light, that is  $0.028c^2$ .

The binding of elements at the different levels of aggregation requires energies of different orders of magnitude. The sum of these energies represents the *energy content* of the body, or the *potential energy of its internal structure*. However, in the study

of mechanical properties only those energies need be considered that are of an order of magnitude that can be affected by mechanical forces and by external conditions relevant in this study, such as temperature. The amount of binding energy of the nuclei or of the electrons in the atom, although large in comparison with the binding energy of the atoms or molecules, is not directly relevant in forming the mechanical properties since it cannot be affected by forces applied to the body. The content of mechanical binding energy, that is, of cohesive energy, is represented essentially by the binding energy of the structural elements, which is of the order of magnitude of the energy that can be produced by mechanical forces, or of heat energy. The material properties of the body as a whole as well as of the elements at the different levels of their aggregation are most concisely expressed in terms of this energy.

In the disintegration processes of material bodies such as in dissolution, fracture, or explosion, the binding energies are "released" at the levels at which the disintegration takes place. In order to produce disintegration of the body, energy must be supplied; in fracture the supplied energy is mechanical, in evaporation the supplied energy is thermal, in explosion the supplied energy is chemical or thermal.

**FORCES.** When a system of *external* forces acts on a material body, a system of reacting forces is produced within the body. The reacting forces which are mobilized in response to the external forces in the course of the transformation of the body from its equilibrium position in the unloaded state to its equilibrium position under the acting forces are usually defined as the *internal* forces. In this sense, gravity, inertia, and magnetic forces are external forces, although their points of application are located within the body; internal forces act between points within the body.

A body may retain its geometrical shape by virtue of the internal forces, or a particular shape may be forced on it by external forces. Bodies that are unable to assume and retain indefinitely a given geometrical shape without the aid of external forces are usually defined as being *liquid*. Conversely, the ability to assume and retain indefinitely any given geometrical shape, independently of external forces, is the characteristic of a *solid*.

The geometrical configuration of the body in the condition of

equilibrium under the acting forces, in particular the configuration of its surface, differs from that of the unloaded body. As a change occurs in the coordinates of the points of application of the external forces, work is done on the body; internal forces produce work only if individual points of the body are displaced with respect to each other. Motion of the body as a whole does not involve work of the internal forces. If in the course of a *virtual* displacement, that is, of an arbitrary variation of the coordinates compatible with geometrical constraints, no work should be done, the work of the external forces must likewise vanish. Thus, a deformed body can be in a state of equilibrium only if the applied external forces are in equilibrium.

This equilibrium condition is valid for any body and for any part of it cut out of the original body. The forces acting across any surface created by a cut are the resultant forces of internal interaction across this surface prior to the cut; in the most general case they can be represented in space by one force and by one moment about the axis of the force.

**DEFORMATION.** The differences between the geometrical coordinates of individual points of the material body in the loaded and in the unloaded condition represent the *displacements*, the entity of which makes up the *deformation* of the body. Since the internal forces represent the reactions associated with the deformation produced in response to the external forces, deformation and internal forces are interdependent and cannot be treated independently. Laws describing this dependence can either be found empirically by observing the relations of applied forces and resulting deformations on material bodies of finite dimensions, or they may be derived theoretically from concepts concerning the internal structure of the material.

A material body may be loaded by external forces only to such an extent that it is able to mobilize reactions (internal forces). If these reactions are not sufficient to support the system of external forces in a state of equilibrium, the deformations will not tend to reach finite values, and, hence, a state of motion or *flow* will ensue. This state cannot be described in terms of forces and deformations alone, since there is no longer a definite interrelation between them. The mechanical behavior must be described in terms of both the rate of deformation or of flow and the deformation. Conditions of flow for which the flow rate

changes from point to point of the medium but does not change with time are called *stationary* or *steady*. If the absolute values of the flow rates are very small, a state of steady motion may be approximated by a sequence of states of equilibrium, thus re-establishing the direct interrelation between forces and deformations.

The deformation of a material body is accompanied by changes in the body which are mechanical, thermal, electrical, magnetic, optical, and chemical. The nonmechanical changes are, however, of interest in the study of mechanical properties only so far as they affect the relations between forces and deformations or indicate changes in those relations. Thus a change in dimensions is produced in certain crystals when they are subjected to an electric field (piezoelectricity) or to a magnetic field (magnetostriction); on the other hand, changes in magnetic permeability or electrical resistivity in certain metals have been correlated with changes in the mechanical damping or with the extent of internal damage to the cohesion produced by repeated loads.

A certain part of the deformation occurs instantaneously, in step with the loading, whereas another part proceeds with a time lag and may even go on after the load has been removed. On removal of the forces a partial reversal of the deformation takes place instantaneously, and a further part of the deformation is gradually recovered; the part remaining permanently after the external forces have been removed represents the truly irrecoverable or permanent deformation.

The total deformation of a material body is made up of the change of its initial volume, that is, the *change of density*, and of the change of its initial shape, that is, the *distortion*. Thus the total deformation can always be split into its *volumetric* and its *distortional* components.

In mechanical tests and in service, differences in the character of the deformation of bodies are usually determined from observations made during the application of the external forces. It should be realized, however, that the characteristic shape of the load-deformation curves for fundamentally different types of materials, such as metals and high polymers, can, by the selection of appropriate testing conditions, be made to look very much alike. Such similarity will hold for the *loading* range only; it is the behavior during the *unloading* of the material and

also after the complete *removal* of the forces that reveals the basic differences in the character of the deformation in materials that show similar behavior during the loading.

Comparison of deformational behavior of different materials under different conditions requires the separate consideration of the volume change and of the distortion. Volume changes may be assumed to be instantaneously recoverable on removal of the acting forces; distortions consist of three different parts which must be considered separately: the instantaneously recoverable, the delayed recoverable, and the irrecoverable. While the forces are being applied, the three parts of the distortion appear as an entity; the difference between them becomes manifest only during and after the removal of the forces. Whereas the instantaneously recoverable part of the deformation is usually designated as the *elastic deformation*, the delayed recoverable and the irrecoverable parts of the deformation are designated as the *inelastic deformation*, although the delayed recoverable deformation is, in fact, a retarded elastic deformation.

The relations between the forces applied to a material body and the resulting deformation, both elastic and inelastic, expressed and interpreted in terms of the basic concepts, and their implications as briefly considered in the foregoing discussion, constitute the subject of the mechanics of general deformable bodies.

### 3. Methods of Approach

**LEVELS OF APPROACH.** The approach to the study of the mechanics of deformable bodies depends on the level of aggregation of the structural elements at which the mechanical behavior is analyzed. The approach may be on each of three levels of aggregation of the elements, namely:

1. The large-scale or *phenomenological* (engineering) level at which the material is considered to be continuous and homogeneous, being made up of identical volume elements of finite dimensions.
2. The *structural* level at which the material is still considered to be continuous but nonhomogeneous, being formed of elements of different properties and of finite (macroscopic or microscopically observable) dimensions, which fill the space continuously.
3. The *atomic* or *molecular* level at which the material is considered to be discontinuous, being made up of discrete particles of atomic or molecular size.

The lowest, atomic or molecular level, at which the element is the atom or the molecule and at which the interaction between atoms and between molecules produces the dominant mechanical effects, extends over a range from atomic and molecular dimensions of less than  $10^{-7}$  cm up to dimensions of  $10^{-4}$  cm. This range of magnitude of structural elements can be experimentally studied by X-ray or electron diffraction, by observation of the infrared spectrum, and of the spectrum of mechanical vibrations, and by direct observation with the electron microscope. Theoretical analysis within this range is based on methods of statistical mechanics and on the analysis of suitable atomic models, which are simplified representations of the assumed atomic structure. It is at the atomic and molecular level of aggregation of elements that all phenomena originate that are responsible for the effects of time and temperature on mechanical properties and for the so-called "structure-sensitive" properties of materials.

The structural level extends from the upper limit of the molecular level up to the phenomenological level. At this level the material is considered to be built up continuously but not homogeneously of elements of dimensions varying from an order of magnitude of  $10^{-4}$  cm to several centimeters or even several inches. These elements are or can be made optically visible; their effect on mechanical properties can be studied by direct observation in mechanical tests and by observation of the spectrum of mechanical vibrations. Theoretically, the effect of this level of aggregation of elements on the mechanical properties can be studied by methods of analysis of mechanical models of the structure and by methods of statistical summation over the behavior of individual elements forming the body. These elements, usually of different size, are considered to be individually homogeneous and isotropic, but are distributed and oriented at random, that is, so that all positions and orientations are equally probable.

There are two aspects of the analysis of mechanical properties at the structural level of approach: the *synthetic* approach attempts to deduce the phenomenological behavior of the structurally complex continuous material from the known behavior of the individual structural elements, whereas the *analytical* approach aims at obtaining a picture of the structure of the

material from observed relations in large-scale tests. In the latter approach a theoretical concept of the structure of the material is used to predict these relations; the accuracy of the prediction, evaluated by comparison with the test results, is assumed to provide an indication of the validity of the structural concept. The analytical approach has been extensively used in the study of structures of high polymers;<sup>3-1</sup> the synthetic approach has been applied in the investigations of the viscosity of suspensions<sup>3-2</sup> and in studies of the work hardening of metals.<sup>3-3</sup>

It is at the structural level that properties originate which are expressions of structural changes produced by the action of forces, such as the hardening of metals under large strain (work-hardening), the "softening" of gels and of high polymers under the action of forces (thixotropy), and the "stiffening" of fibrous materials under large strains (strain anisotropy).

In the analysis of the mechanical behavior of material bodies at the phenomenological level the material is considered to consist of elements of macroscopic dimensions which have identical properties and are large enough to include all types of behavior in the average ratio. After the deformation of the body has occurred, initially adjacent elements remain adjacent. The mechanical behavior of a body can be determined from the behavior of the element by simple integration over space. The material is considered to be *isotropic* if, for every element, the various directions in space are mechanically equivalent. When, however, the continuous medium representing the material, or a finite part of it, cannot be mechanically described without its orientation being specified, the material is defined as *anisotropic*.

The introduction of the concepts of homogeneity and isotropy in relation to real materials can only be justified on a statistical basis, considering the average shape and material properties of the elements making up the body. It is evident that in the case of such statistical homogeneity and isotropy the relations between the quantities describing average behavior and properties of the body and those describing individual behavior and properties of the constituent elements can only be statistical relations. There is no more reason to assume a definite functional relation between the average mechanical properties of the large group of elements forming the body and the properties of any individual constituent element, than there would be to assume such

a relation between the life expectation of men in a certain age group specified on the basis of a statistical mortality table and the actual length of life of any particular individual in this group.

**RHEOLOGICAL THEORY.** The assumption of statistical homogeneity and isotropy of the continuous material body makes it possible to develop a general phenomenological theory of deformational behavior of materials. This theory forms an organic part of the mechanics of deformable media and provides the transition between the classical theory of elasticity and classical hydrodynamics, both of which are included as limiting cases of the general theory of *rheology* (science of flow).

Although the deformation of real materials is a rather complex phenomenon, it has been observed that similar processes of deformation can be produced in different materials by varying the intensity of the load, its character, its rate of application, the temperature and the shape and dimensions of the loaded body. On this basis it has been assumed that, phenomenologically, mechanical behavior of real materials is governed by a few characteristics only which, in different combinations, are present in all materials. On this assumption the mechanical response to external forces and temperature would be determined by the relative importance of the constituent *phases*, each of which possesses certain of the basic characteristics. The term "*phase*" as used here refers to mechanical behavior only and defines different parts of the material characterized by different types of response to imposed forces. As the medium is considered to be homogeneous, the different phases are not necessarily identifiable with different parts of the volume but represent rather the different types of mechanical response present in each volume element.

Thus, an effective quantitative theory of deformation under external forces can be established, provided that the relations governing the deformational response of each of the constituent phases can be formulated in relatively simple mathematical terms, accessible to analytical treatment. These relations are generally expressed as functions of the variables specifying the *dynamical* conditions of the considered volume element (forces, stresses) and the variables specifying its *kinematical* conditions (displacements, strains). They can be formulated on the basis of the assumptions that within the homogeneous deformable medium the dynamical and the kinematical variables and

their derivatives are continuous and that the behavior of the medium at a certain point depends on the conditions at this point only, being independent of the conditions in the vicinity of the considered point, that is independent of the *gradient* of stress or strain. This latter assumption is one of expediency; its justification has not been proved. Without this assumption, however, a mathematical approach to problems of mechanical behavior would be practically impossible.

The establishment of relatively simple mathematical relations describing the mechanical behavior of real materials requires a certain idealization of actual behavior. Classical mechanics of deformable materials, which has, so far, provided the basis of practically all engineering analysis concerned with strength and deformation, has been developed from two idealizations, namely that of the linear elastic or Hookean solid and that of the ideal viscous or Newtonian liquid; both materials are defined by a linear relation between the relevant mechanical variables. Originally these relations were introduced not as idealizations of real behavior but as laws expressing the deformational behavior of real materials, and they were accepted as such for a long time. However, the limitations of the classical theories of deformable media were increasingly realized as the consideration of mechanical behavior beyond the range previously considered as "practical" (within which the approximation provided by the classical theories is very close) became more important. Consequently, the need has been felt for a broadening of the underlying assumptions concerning the idealized types of deformational behavior which would lead to the development of a general theory of mechanical behavior of real materials. Although it was recognized that only the idealized materials are limited in their properties, because this is how they are defined, it was also found that idealized materials can be combined so as to represent real materials by different degrees of approximation.<sup>3-4</sup> Rheology constitutes an attempt to make use of this fact and to establish a phenomenological theory of general behavior of materials on the assumption that every real material possesses all basic deformational or, as the terminology is, all *rheological* properties in varying proportions.

MECHANICAL BEHAVIOR OF GROUPS OF ELEMENTS. In studying mechanical properties of real materials the most important con-

sideration is that such materials are neither continuous nor homogeneous or isotropic. Because they are built up of large numbers of discrete elements or particles, the observed large-scale properties are actually the result of the behavior of the aggregation or of the *group* of elements or particles that make up the material. Hence, the relation between the forces applied to the material body and its deformational response might be derived from the consideration of the response to these imposed forces of the discontinuous group structure of the body, thus expressing or *interpreting* the large-scale mechanical properties in terms of the group behavior of the elements or particles. In this manner the observable phenomena are deduced from non-observable processes, and the phenomenological laws are derived from more fundamental microphysical laws. This approach to problems of mechanical behavior leads to atomic or molecular theories of such behavior.

The study of mechanical properties means the study of deformational behavior; it is through behavior that properties manifest themselves. Group behavior is made up of the characteristic behavior of the individual elements as well as of their behavior in the group, which is determined by their mutual interaction. The group properties are therefore not identical with the sum of the properties of the individual elements. The group has an individuality of its own which is determined by the laws according to which it has been formed. The aggregation of elements thus produces a certain *group pattern* by which the group can be identified. For example, in the aggregation of atoms forming a crystal or in the aggregation of molecules forming a high polymer, the observable deformational response of the material body to imposed forces is made up of the responses of the atoms or molecules integrated into a group by the binding forces between them. In forming the total response, the laws according to which the binding forces between the particles are established and are changed with changing external conditions (and which thus represent the laws of group behavior) are considerably more important than the behavior and the intrinsic properties of the individual particles.

Some properties of the material, however, are not affected by the group pattern. For example, the weight of the body is practically unaffected by the fact that the particles are bound;

it is simply the sum of the weights of the individual particles. In comparison with the total mass of the body, the "mass defect" related to the binding energy (see Art. 2) is negligibly small.

Thus, two types of properties must be distinguished: properties that are not affected by the group pattern, usually termed *additive*, and properties essentially determined by the group pattern, usually termed *constitutive*.

In considering a material body as made up of groups of elements aggregated at different levels of organization, a certain level must be selected at which the concept of a group structure formed of discrete elements is replaced by the concept of the continuous medium. The concept of continuity is usually introduced at the level of aggregation at which the number of group elements that are simultaneously affected or simultaneously observed is so large as to make the individual contribution of an element to the behavior of the group unidentifiable. When this condition exists the concept of the statistical average effect, functionally unrelated to the effect produced by the individual element, must be introduced.

Fundamentally, all physical processes and distributions are discontinuous. The appearance of continuity in space and time is created by the large number of individual discontinuous processes making up the phenomenon, as well as by the relation between the duration of the process and the duration of the observation. Apparent continuity of a process is therefore the effect of a space and time integration over the basically discontinuous processes. For example, a metal crystal that appears to be continuous produces a relation between force and deformation that for slow rates of deformation is markedly discontinuous. The slower the process and the closer the observation, the more discontinuous the observed deformation. On the other hand, the more rapid the deformation, the more continuous will it appear. The deformation of a polycrystalline metal appears to be continuous, even under close observation because of the large number of crystals involved, although the deformation is made up of a number of sharply discontinuous slip motions. The finer the grain, that is, the larger the number of crystals involved in the process, the more continuous the observed deformation.

Since the groups of elements at one level of aggregation represent the individual elements at the next higher level, the laws of

group formation govern the mechanical behavior of materials and of structures at all levels of aggregation. The level may vary from that at which, for example, iron and carbon atoms are grouped into the geometrical crystal pattern to that at which polycrystalline steel shapes, assembled into structural members, are associated to form the geometrical pattern characteristic of a certain bridge truss. At all levels of aggregation the mechanical properties are determined essentially by the group pattern.

The group pattern has a dual aspect: the geometrical and the material. The geometrical pattern defines the arrangement in space of the elements or particles of the group and the group boundaries, whereas the material pattern, being an expression of the distribution of the mass and of the "mass defect" of the particles, is directly related to the binding or potential energy of the group. It represents the distribution over the particles of the potential of the interacting forces, that is, of the energy expended in forming the group out of the isolated elements. The combination of the geometrical and the material pattern or energy pattern determines the mechanical response of the group, at any given moment, that is, its *mechanical state*.

The geometrical and the energy pattern of a group of elements are not independent; in fact, the condition of stability, formulated in terms of a condition of minimum energy of the binding forces, determines the geometrical pattern of the group. Mechanical behavior is defined by the change of state of the group, that is, by the change of both the geometrical and the energy pattern.

The larger the number of elements forming the group at a certain level, the less is the group behavior affected by the properties of the individual element, and the more pronounced is the influence of the laws of group formation. Thus the mechanical behavior of a material body consisting of a few large crystals can, with a certain approximation, be deduced from the observed behavior of a single crystal. However, neither the single crystal nor the body consisting of a few large crystals can be expected to provide the basis for the study of the behavior of polycrystalline aggregates, since in such aggregates the group behavior, that is, the interaction between the crystals, is at least as effective in influencing the behavior of the aggregate as is the behavior of the single crystal.

The considerable importance of chance fluctuations in the

observed large-scale mechanical properties of real materials is the result of the discontinuous internal structure of such materials. Since the mechanical properties of a material on a phenomenological scale are determined by the behavior of the groups of elements forming it at the different levels of aggregation, the chance effects, which strongly influence the process of aggregation of elements, must necessarily have an influence on large-scale behavior.

A material property which is the result of the combined action of a number of elements will, according to statistical principles, be the more uniform, the larger the number of elements involved in its creation. Therefore, if a group property depends on the behavior of a small number of selected elements in the group, the fluctuation of the values of this property will necessarily be considerably wider than that of a property that depends on the interaction of all or of a large number of elements. This consideration explains the observed difference in the range of fluctuation of the values of so-called *structure-sensitive* and of *structure-insensitive* properties. The structure-sensitive properties, such as, for example, the fracture strength, are essentially determined by local imperfections in the group structure and, therefore, depend only on a small group of elements affected by these imperfections; consequently they show a considerably wider range of chance fluctuation than the structure-insensitive properties, such as, for example, the elastic constants. Whereas the structure-sensitive properties depend on the *selective* contribution of a small group of anomalous or anomalously located elements within the large group, structure-insensitive properties are the result of the *additive* contribution of all elements to the average behavior of the group.

Very few materials are built up from a single sequence of groups. Frequently two or more sequences of different group structure are combined "in parallel" to form the so-called *skeleton structures*. The simplest structures of this type consist of a spongelike continuous solid skeleton of one material the voids of which are filled by a different, usually viscous material or of a fiber-like continuous skeleton of one material embedded in a large volume of a different material. Because of the parallel group formation that has taken place, the mechanical behavior of such materials is the result of the interaction of the two groups

of elements. Since groups characterized by different response to the imposed forces have been termed *phases*, materials of skeleton structure are frequently designated as *two-phase* or *polyphase* materials, according to the number of the different materials that have been combined. The concept of the *two-phase* structure in which one phase is solid and the other behaves essentially as a liquid of high viscosity is used extensively in later chapters dealing with the structure of technically important materials.

**ENERGY CONSIDERATION.** The study of mechanical behavior and properties at the three different levels of the aggregation of elements that compose the material, namely, the phenomenological, the structural, and the atomic or molecular, requires the consideration simultaneously and the correlation of the phenomenon on the different levels in order to develop a general integrated theory of mechanical behavior.

On the phenomenological level the mechanical behavior of the continuous body is described in terms of the relations between stresses and strains and their derivatives. On the atomic or molecular level the behavior of the discrete particles is described by their relative positions in space, their velocities, and the forces of interaction between them. No correlation is possible between the stresses and strains in the volume element on the one hand and the atomic or molecular forces of interaction and the change of relative position in space of particles on the other. Correlation of behavior on the different levels is possible only in terms of a concept which on all levels has the same meaning in both Newtonian and statistical mechanics, the same dimension, and the same tensorial rank.

This concept is *energy*. Being a scalar, that is, a tensor of rank zero, it is an algebraically additive quantity and has the same meaning on all levels of group formation. Hence, a description of mechanical behavior in terms of energy provides a method by which the phenomena at the different levels of group formation can be described in identical terms and correlated by the correlation of the processes of energy transformation at the respective levels. If mechanical and heat energy only are considered as relevant in the study of mechanical behavior, the laws of energy transformation are expressed by the first and second law of thermodynamics. These two laws provide, therefore, the

basis for the analysis of the mechanical behavior of materials at all levels of aggregation of elements.

The basic laws of thermodynamics have been formulated both in terms of classical (phenomenological) and of statistical mechanics and are therefore equally applicable in the analysis of the behavior of discontinuous systems containing large numbers of particles and in the analysis of the deformable continuous medium. This interrelation between classical thermodynamics and statistical mechanics provides the methods by which the laws of mechanical behavior at the different levels of investigation can be correlated. Since the interrelation has been derived from the consideration that all conclusions of classical thermodynamics can themselves be obtained from statistical mechanics, the properties of matter, which enter the equations of classical thermodynamics merely as empirical constants to be determined by observation and measurement on a macroscopic scale, can frequently be derived from or interpreted in terms of atomic or molecular processes.

#### 4. Elastic and Inelastic Behavior

In accordance with previous definitions, the mechanical behavior of a material body subjected to forces is called elastic if the deformations produced in the body are instantaneously recovered when the forces are removed. This is the meaning of the usual definition of elasticity as the existence of a one-valued time-independent relation between forces and deformations. The simplest form of this relation is Hooke's "law" of linear elasticity.

The mathematical theory of elasticity developed on the basis of Hooke's law has provided the foundation for the procedures of engineering design. By the introduction of the linear relation between forces and deformations and the restriction to infinitesimal deformations, elastic theory has been linearized and numerical analysis of the elastic response of a continuous homogeneous medium to applied forces has been made practically possible. Because of the linearization of the basic differential equations, the resultant effects of imposed individual conditions can be obtained by algebraic summation (superposition) of the individual effects.

The linear elastic body is a good model of the mechanical

behavior of engineering materials within a practically very important range. If this were not so, attempts to determine or to predict the behavior of engineering structures, even under service conditions, would be prohibitive because of the complexity of mathematical analysis of nonlinear behavior.

Linear elasticity need not necessarily be considered an empirical property. As is shown in a later section, it represents, with very close approximation, the theoretically required behavior under load of any continuous solid homogeneous isotropic body of constant density of whatever material within the range of small (not only infinitesimal) deformations. The deviation of the behavior of real materials from the linear relation between forces and deformations within the range of small deformations is due only to the fact that the concept of the homogeneous medium is but an approximation of the real material and that for certain materials the influence of time may become important.

The level of the elastic limit of behavior in terms of the applied force is a function of the accuracy of its observation; in fact, for an infinitely accurate observation this limit for real materials would be at zero force since some irreversibility of the deformation process must always be expected in accordance with the second law of thermodynamics, although the deviation from elastic behavior may not be perceptible on the scale of observation or may be hidden because of the method of observation. Thus, for instance, for a certain metal observation of the force-deformation diagram in simple tension may show perfect elasticity, whereas definite evidence of inelastic behavior within the same range of forces can be obtained from a vibration test. It is, however, this deviation from elasticity that is of principal importance in determining the significant properties that make up the structural performance under load of engineering materials. Elasticity or near elasticity of a material is but the condition for the application of comparatively simple design methods; it is its *inelasticity*, however, which determines its mechanical properties as well as the range of its use and its safety in service.

The engineer usually considers elasticity of his materials to be the most important property on which to base his design for service. He realizes, however, that the actual performance of the structure might be very different from the assumed performance of the designed elastic structure, and that both the

differences and similarities in the behavior of the real and of the designed structures, as well as the safety of the real structure, are essentially results of the deviation of the behavior of real materials from the linear elasticity assumed in design. Hence, the interest of the engineer necessarily centers on the deviation from conditions instinctively considered "normal," that is, elastic, rather than on generally defined conditions of *rheological* behavior. This deviation from elastic conditions may be considerable, but its importance is still mainly in its relation to an elastic component. Thus, the term *inelastic* behavior or *inelasticity* is meant to define behavior "beyond the elastic range," the implication being that the elastic range remains of considerable if not of primary importance.

The use of the term *plasticity* as synonymous with inelasticity is considered undesirable, because this term is used to define a specific type of "inelasticity" which is characteristic of metals in the range of large deformations; its simultaneous use in the general sense is confusing. Recently Zener<sup>1</sup> has coined the term *anelasticity* to define another specific type of general inelastic behavior, a type that is characteristic of two-phase or polyphase materials in the range of small deformations and which before has often been referred to as *after-effect*. Both *plasticity* and *anelasticity* are subsequently used in the limited meaning referred to, whereas *inelasticity* is applied to designate any type of mechanical behavior that is not elastic.

Elasticity and inelasticity of a one-phase material can be defined with reference to the changes in the geometrical and energy patterns produced in the elements of the material body by applied forces. The body is elastic if all changes that have been produced are instantaneously recovered on removal of the forces. If only part of the change can be recovered by removal of the forces whereas the remaining part is irrecoverable, the behavior is inelastic. The intermediate type of inelasticity defined by delayed recovery of the deformation is characteristic of the two- or polyphase material only; it does not occur in a material consisting of a single phase.

Changes of mechanical behavior may be produced by changing primarily the geometrical pattern with unidentifiable changes of the energy pattern or by changing primarily the energy pattern of the body without identifiable change of the geometrical pat-

tern. Because of the interrelation between geometrical changes and changes of energy distribution there will always be a dual effect, whatever the character of the primary change. However, a classification of inelasticity may be based on the assumption that the character of the mechanical behavior depends on whether the produced changes are primarily in the geometrical pattern or primarily in the energy pattern.

It may reasonably be expected that changes of geometrical pattern are important only when the existing pattern is *unique*, that is, identifiable, having a considerably higher stability of existence than any alternative pattern. Such stability would require that the probability of forming the existing pattern be appreciably higher than the probability of forming any alternative pattern. If, however, a number of patterns of equal stability, that is, of equal probability of existence, can be formed, the existing pattern is no longer *unique*, as it is not recognizably different from any of the alternative patterns. In this case changes in the energy pattern will be the dominant effect governing mechanical behavior.

Perfect reversibility of the mechanical response on removal of the acting forces is possible only if the initial geometrical pattern is identifiable. If a large number of equally probable patterns can be formed, the probability that, on removal of the forces, the initial pattern, which is only one of the possible equivalent patterns, will be exactly re-established is relatively small. Thus, perfect reversibility of deformation is a characteristic of materials whose geometrical pattern is identifiable. Because of the very large number of particles involved, a pattern can be identified only by a definite geometrical periodicity, such as exists in the crystal structure.

Only materials with an identifiable unique geometric pattern, such as crystals, can be truly elastic; thus the energy pattern is of less importance than the geometric pattern in determining the deformational behavior of the *ordered* material. On the other hand, in the *unordered* material many geometrical patterns can be associated with a given energy content; hence, the change in energy toward a lower value can proceed along a number of alternative paths and therefore becomes the dominant factor in controlling the deformational behavior; the geometric pattern has freedom to adjust itself to the energy pattern. This fact

explains why *unordered* (so-called amorphous) material is more sensitive to temperature and time than is *ordered* (crystalline) material.

In most engineering materials inelasticity of deformation is observable under the action of very small loads if adequate testing conditions are selected; it becomes more pronounced with increasing load intensity. By imposing certain loading conditions, the inelasticity of behavior can be intensified so strongly that the deformation of a material usually considered brittle may resemble liquid flow of high viscosity. Therefore classification of materials according to their apparent deformational behavior as ductile, tough, or brittle has no basic physical meaning; it may be useful only as an arbitrary specification limited to well-defined conditions. The material is not invariably brittle or ductile; "brittleness" or "ductility" describe but its momentary behavior, that is, its momentary response to a certain system of forces and conditions. Only as long as a material is assumed to be elastic can its mechanical behavior or response be considered invariable, because elasticity expresses the fact of invariability of mechanical behavior. It is the variability, with respect to the imposed forces, of mechanical behavior and properties that is the mark of inelasticity.

Changes of the internal structure of materials resulting in inelastic behavior originate at the atomic level; an analysis of such behavior must therefore start from the consideration of the relevant phenomena taking place in the atomic structure of the material. The first part of this volume, Chapters 1-3, contains a discussion of the atomic constitution of matter and of the changes in atomic structure produced by external forces and temperature. The discussion of the basic physical concepts is somewhat more extensive than would be absolutely necessary for the understanding of the following chapters; however, it is considered that a certain familiarity with modern concepts of the structure of matter may be helpful to the engineer interested in materials in furnishing the background for independent thought and for the study of advanced books in the field of the theory of solids.<sup>4-2</sup>

The second part of the book, Chapters 4-12, is an attempt to present an integrated theory of inelasticity of which the theories of elasticity, of plasticity, of work hardening, of creep and

relaxation, and of anelasticity constitute special cases. This general theory, although primarily phenomenological, is based on the structural concepts developed in the first part of the book. Chapter 11 contains a discussion of fracture as the terminal point of deformation. A short discussion of the behavior of suspensions and gels is included in Chapter 12. These materials, of primary importance to the chemical engineer, may not be of direct interest to the engineer concerned with load-carrying materials, unless he is interested in soil mechanics. However, an adequate understanding of the process of formation of solid materials requires some knowledge about the behavior of these semiliquids out of which many solids are formed.

The last part of the book, Chapters 13-18, deals with the engineering applications of the theory of inelasticity, in particular with special problems of the theories of plastic equilibrium and of plastic flow, as well as of the theory of work hardening. Chapters 16 and 17 contain a discussion of the principles of design for plasticity, for work hardening, for creep, and for fracture. The analysis of the engineering significance of inelasticity is concluded (Chapter 18) by a discussion of the significance of mechanical testing of engineering materials.

### References

- 1·1 M. REINER and R. SCHOENFELD-REINER, *Kolloid-Z.* **65** (1933) 44.
- 2·1 M. BORN, *Atomic Physics*, Blackie & Son, London (1945) 61.
- 3·1 E. GUTH and H. M. JAMES, *Ind. Eng. Chem.* **33** (1941) 624; *Phys. Rev.* **59** (1941) 111.
- 3·2 G. B. JEFFREY, *Proc. Roy. Soc. A* **102** (1923) 161.
- 3·3 A. M. FREUDENTHAL and M. REINER, *J. Applied Mechanics* **15** (1948) 265.
- 3·4 M. REINER, *Rheol. Bull.* **16** (1945) 53.
- 4·1 C. ZENER, *Elasticity and Anelasticity of Metals*, Univ. of Chicago Press (1948).
- 4·2 F. O. RICE and E. TELLER, *The Structure of Matter*, John Wiley & Sons, New York (1949).

P A R T

A

The Structural Aspect  
of Mechanical Behavior



# CHAPTER

## 1

### GENERAL CONCEPTS OF THE STRUCTURE OF MATTER

#### 5. Oscillators and Quanta

OSCILLATORS: CONCEPT OF ATOMIC STRUCTURE. The physical concepts by the aid of which the atomic structure of matter is described are based either on an admittedly fictitious, pictorial model of atoms and of atomic processes or on a realization that there is no possibility of pictorial representation of atomic happenings in terms of space, time, and causality and that, therefore, only abstract mathematical reasoning is adequate to describe what is assumed to take place on the atomic level.

The study of the atomic structure of matter started from a rather crude pictorial model concept. In the course of time, this concept became increasingly complex, as it had to fit all the newly discovered experimental facts, until its inadequacy became so glaring that it could no longer be saved by further complicating refinement. At this stage in the development, attempts to visualize atomic processes were abandoned in favor of the non-pictorial approach of quantum and wave mechanics which proved to be remarkably powerful and incisive. In spite of the complexity of the mathematical tools required for its application, this approach leads to an explanation of the basic nature of atomic phenomena which is of surprising universality and fundamental simplicity.

There is no doubt that the use of pictorial models will remain an important and legitimate approach in the interpretation of

nonperceptible physical processes. However, the interpretation of the meaning of even the familiar looking model may require a redefinition of the model elements used.

The simplest pictorial concepts of atomic structure of matter does not attempt to introduce any model of the structure of the atom itself. The atom visualized as a minute (rigid or elastic) sphere of matter is considered to be the elementary particle, subject to a combination of attractive and repulsive forces which determine its position within the group of surrounding atoms. The structure of the atom is not considered relevant at this level of the investigation. Each atom is thus visualized as an "oscillator,"

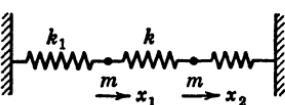


FIG. 5·1 Two-particle oscillator model.

that is, a mass point which, following the laws of Newtonian mechanics, oscillates about its equilibrium position, in which it is "elastically" held by the electrostatic forces exerted by the surrounding atoms. Its total energy is made up of the

energy of the electrostatic potential and of the kinetic energy of its oscillating mass.

The oscillator model of matter is not a conservative system, since in the process of oscillation particles are able to absorb or emit heat energy by oscillating more or less violently, that is, with a larger or smaller amplitude. The distribution of the binding or potential energy of the individual particles of the material centers around the mean potential energy. With a comparatively small number of particles having potentials of the binding forces considerably higher or considerably lower than the mean value, the number of particles possessing a certain potential or *energy content* is assumed to decrease with increasing deviation of this energy from the mean value. Before the advent of the quantum concept, it was assumed that within the large number of oscillating particles forming the material body all frequencies of oscillation could exist.

**TWO-OSCILLATOR MODEL.** The behavior of an oscillator can, in rough approximation, be made clear by the analysis of the simple mechanical model of a linear oscillating system with two degrees of freedom shown in Fig. 5·1, in which the two equal masses  $m$  represent the particles bound to each other and the surrounding particles by electrostatic forces (represented by

springs). This model is based on the simplifying assumption that only neighboring particles interact.

The most general motion of the system is made up of the superposition of two harmonic oscillations in which the two masses oscillate with the same frequency and in the same phase.<sup>5-1</sup> The lower frequency represents the *fundamental* or *symmetric* mode, in which the two masses swing in the same direction with equal amplitudes (the coupling spring is not stressed). If such motion involves a line of particles in a crystal lattice, this lattice oscillates as an entity, that is, as if it were part of a continuous body. Such translatory motion of a line of particles is therefore associated with the propagation of the acoustic (elastic) vibrations of the continuous elastic medium. The higher frequency represents the *antisymmetric* mode, in which the two masses move against each other with equal amplitudes. Motion of this type along a line of particles in a crystal lattice is the effect of different (positive and negative) electric charges of adjacent particles; it is thus typical for electrodynamic (optical) oscillations of the lattice.

The two modes of oscillation are not coupled; each of the motions can take place independently. Therefore, the behavior of a group of particles bound together into a lattice can be represented by the behavior of a group of uncoupled oscillators.

Equations of *beats* result from the superposition of the two principal oscillations.<sup>5-2</sup> Each mass executes a rapid vibration with an amplitude that changes slowly. The masses move in opposite phases, so that the amplitude of the oscillation of one mass increases while the other decreases, reaching its maximum while the other is at rest. The beats are slow when the coupling is weak and become rapid as the strength of the coupling increases. The process represents an exchange of energy between the two degrees of freedom and is of the same type as the well-known phenomenon taking place between two tuning forks on a common base after one of them has been excited.

The natural frequencies of the oscillators are identical with the frequencies of radiation at which energy is absorbed or emitted by them, as these frequencies may be considered the resonant frequencies with regard to the frequency of the emitted or the absorbed radiation.

The natural frequencies of the oscillation of the particle within

the group do not represent the complete spectrum of frequencies of the individual oscillator. They only represent the frequencies associated with the interatomic or intermolecular bonds; the energy of the oscillators can be expressed as the *heat content* or the heat energy of the body since, as is shown later (Art. 10) the wave length of the radiation emitted or absorbed at the natural frequencies of the oscillations of the bound particles is in the *infrared* range. This is the radiation that has the temperature of the object that emits it.

A number of natural frequencies of much higher order exist at which oscillators will emit or absorb radiation of light or radiation of still higher frequency, such as X-ray or nuclear radiation. These frequencies are, however, not directly related to the mechanical behavior of the material formed by the group of oscillators; only the infrared (thermal) radiation is associated with the behavior of the group.

QUANTA. Studying the energy distribution of the infrared radiation, Planck concluded that the observed facts of radiation could be explained only by assuming that, within the large number of oscillators forming the material body, not all natural frequencies of oscillation or frequencies of radiation can exist, since energy is not emitted or absorbed continuously, but is contained in the oscillating particles only in whole multiples of energy units or *energy quanta*.<sup>5-3</sup> The units themselves vary with the natural frequency of oscillation of the particles; they are large for small wave lengths of radiation or large frequency and small for large wave length or small frequency.

According to this *quantum* concept, the Planck oscillator (unlike the previously considered Newtonian oscillator) can neither absorb nor emit heat or any other radiation energy continuously. The acquisition or emission of energy can only take place in discrete amounts or bundles of energy, called *quanta*; it is therefore a discontinuous process. Thus, if heat is applied to a material body consisting of oscillators, it cannot uniformly increase the energy, that is, the amplitude of the particles oscillating about their equilibrium position; it can only increase the energy of the individual particle by an integral multiple of the discrete amount of energy, the *energy quantum*, which is a characteristic of the particle, or rather of the frequency at which it oscillates. The core of Planck's quantum theory is the state-

ment, that the energy quantum  $\epsilon$  is related to the frequency  $\nu$  of the oscillations of the particle by the equation,

$$\epsilon = h\nu \quad (5 \cdot 1)$$

where  $h = 6.62 \times 10^{-27}$  erg-sec is a universal constant, known as the *unit of action* or Planck's constant.

Since the vibrational energy of the oscillating particle is absorbed or emitted only in whole numbers of energy quanta, this number is related to the radiation frequency. If  $E_1$  and  $E_2$  denote two energy states of the oscillator, the relation holds

$$E_1 - E_2 = \Delta E = \Delta n h\nu \quad (5 \cdot 2)$$

where  $\Delta E$  is the difference in the energy content of the two states and  $\nu$  denotes the frequency of the emitted ( $E_1 > E_2$ ) or the absorbed ( $E_1 < E_2$ ) radiation. Hence

$$\nu = \frac{1}{h} \frac{\Delta E}{\Delta n} = \frac{1}{h} \epsilon \quad (5 \cdot 3)$$

where  $\Delta n$ , denoting the difference in quantum numbers between  $E_1$  and  $E_2$ , can only have discrete values, as  $\Delta E$  is always an integral multiple of  $\epsilon$ .

This conclusion is borne out by the existence of a certain limited number of *spectral lines* in the light spectrum of various chemical elements which had been observed long before the advent of quantum theory. However, quantum theory explained why the spectral lines, which are important identifying marks of chemical elements and represent frequencies of absorption of radiation energy, appear only at certain frequencies; these are the differences between the consecutive frequencies of oscillation in the "quantum ladder" of frequencies defined by the consecutive energy levels, which are whole multiples of the quantum.

According to Planck's quantum relation, absorption or emission of energy by particles oscillating at high frequencies proceeds by larger energy bundles than the bundles needed for absorption or emission of energy by particles oscillating at low frequencies. However, the higher the mean energy of the particles, that is, the larger the average number of energy quanta they hold in oscillating, the smaller the relative difference between consecutive energy levels and the less, therefore, the difference between a continuous (classical) distribution of energy levels and the

actual quantum distribution. This is only a different expression of the fact that at a mean energy level of a few quanta the jump by one quantum will be much less probable and therefore much more difficult than the same jump at a mean energy of many thousands of quanta, although the absolute step is identical, since this energy quantum will become the more easily available, the larger the mean energy. Being thus able to change its energy level with comparative ease, the particle in states of high mean energy, which are states of relatively high temperature, becomes increasingly "mobile" and less selective with regard to the particular energy level it is willing to occupy. Thus, at a certain temperature the quantum-mechanical concept of the Planck oscillator approaches the classical concept of the Newtonian oscillator and the quantum-mechanical laws approach the laws of classical mechanics.<sup>5-4</sup>

This same consideration explains why the quantum concept loses its importance, though not its validity, in the description of the behavior of macroscopic bodies. If the energy associated with a process of a few seconds duration in which a macroscopic body is involved (for instance, a small swinging pendulum or a small piece of heated metal) is of the order of magnitude of a foot-pound or a few calories, that is, of an order of magnitude of  $10^6$  to  $10^8$  erg or, in units of action,  $10^6$  to  $10^8$  erg-sec, a discontinuity in the energy change expressed in steps of an order of magnitude of  $\hbar\nu = 6.6 \times 10^{-27}\nu$  or about  $10^{-14}$  erg (if the frequency of heat radiation  $\nu = 10^{12}$ ) can have no significance; therefore, the process appears perfectly continuous. It is only when the energy associated with the process is of the order of magnitude of a few quanta that the inherent discontinuity manifests itself.

When the particle passes from one quantum (energy) state to another, the laws of conservation of energy and of momentum must hold. This conclusion is based on the statistical consideration that the macroscopic behavior of the group of particles, for which the conservation laws are known to hold, requires that these laws be true, at least statistically, for the individual particle. As long as no particular mechanism of the transition from one quantum state to another is conceived, conclusions concerning the frequency of such transition can only be derived from the probabilities of spontaneous emission of a certain energy by a particle under a given set of external conditions. The frequency

of transition of a particle from a higher to a lower energy level depends on the number of existing alternative states of lower energy to which the particle in the *excited* state (that is, in a state of energy above the stable mean energy level) can jump spontaneously, emitting the energy difference. If  $p_{ji}$  denotes the probability per second of a spontaneous change from the quantum state  $j$  to the lower state  $i$ , the "mean life" of a particle in the excited state is  $\tau = 1/p_{ji}$  sec. In case there is not one, but a number of possible lower states  $i$ , the "mean life" is reduced to  $\tau = 1/\Sigma p_{ji}$ . Thus, the higher the momentary quantum state of the particle above that of mean energy, that is, the larger the number of intermediate energy or quantum levels to which the particle can jump, the shorter the "mean life" of the particle in the excited state and the higher therefore its rate of transition to a lower state.

## 6. Bohr's Model of the Atom

THE "SOLAR-SYSTEM" MODEL. The reason for the introduction of the atomic model is the necessity of explaining and accounting for the existence of interatomic bonds. With the aid of the atomic model the forces of interaction between atoms can be related to the energy contained in the structure of the atoms which, according to the theory of matter, is electric energy. The consideration of the structure of the atom and of the real nature of the interatomic bond is not only of theoretical but also of immediate practical interest in the study of large-scale mechanical behavior. All large-scale fracture phenomena, for instance, are initiated on the atomic or molecular scale and are essentially governed by atomic or molecular processes. Unless the nature of the atomic bond and the conditions for its formation and disruption are understood, fracture phenomena cannot be generally interpreted and explained.

Pictorial concepts of the structure of the atom are based on the "solar system" atomic model proposed by Rutherford and perfected by Bohr.<sup>6-1</sup> Rutherford, realizing the electric nature of matter, showed that in the atom the whole positive charge and almost all the mass must be concentrated in a single compact heavy nucleus, surrounded in some way by moving negatively charged particles, the electrons, the combined negative charges of which just balance the positive charge of the nucleus. The

whole system thus resembles a minute solar system; its operative forces are the electrostatic attraction between charges of opposite sign and the centrifugal forces of the electron, resulting from its mass and acceleration.

That Rutherford's atomic model, which is the basis of all modern ideas of the structure of the atom, was inadequate, became evident when it was realized that a Rutherford atom would emit energy of all frequencies; such behavior could not reproduce the observed definite spectral lines. It would, moreover, emit this radiation at a rate considerably higher than the rate observed and therefore disintegrate rapidly.

The Rutherford model was perfected by Bohr who applied Planck's quantum concept to the "planetary" motion of the electrons. Bohr's two conditions by which the unstable Rutherford model is transformed into the perfectly stable Bohr model require that

1. Among the infinite number of possible orbits, the electrons rotate continuously, that is, without the emission of energy, only within certain orbits designated as *stationary state orbits*; these are elliptical with the nucleus at the focus of the ellipse.

2. The stationary orbits are defined by the condition that electrons moving along them may possess only integral multiples of energy quanta  $h\nu$  where  $\nu$  denotes the frequency of rotation.

The angular momentum of an electron of mass  $m$  in a stationary state of circular motion, which is the simplest type of orbital motion, is obtained by dividing the energy of the rotating electron by  $2\pi\nu$  and is therefore an integral multiple of  $h/2\pi$ , that is,  $nh/2\pi$ . By passing from one stationary energy orbit to another, the electron absorbs or emits a number of energy quanta equal to the difference of the quantum states of the electron in the respective orbits.

The analogy of the solar system and of Bohr's atomic model breaks down at the point where the characteristic discreteness of the stationary orbits and of the associated quantum states in the Bohr model is realized.

According to the laws of electrostatics the attractive force between charges  $e$  of opposite sign is  $(e^2/r^2)$ . Hence the centrifugal force  $mv^2/r$  required to balance an electron attracted by the nucleus in its orbit of radius  $r$  is obtained from the relation  $mv^2/r = e^2/r^2$ ; hence,  $mv^2 = e^2/r$ . Since, according to this rela-

tion, the velocity  $v$  of the electron decreases with increasing radius of the orbit, its kinetic energy decreases with increasing distance from the nucleus, whereas, its potential energy increases; at an infinite distance from the nucleus the potential energy is necessarily zero. At the same time, it must reach there a maximum as the work done by the forces of attraction increases with increasing radius of the orbit; the potential energy of the electron for all finite distances from the nucleus  $r > 0$  must therefore be negative. Hence the total, that is, potential and kinetic energy  $E = \frac{1}{2}mv^2 - e^2/r$ . Because of the equilibrium condition  $mv^2 = e^2/r$ , the energy  $E = -\frac{1}{2}mv^2 = -\frac{1}{2}e^2/r$ .

PERIODIC TABLE OF ELEMENTS. PAULI'S EXCLUSION PRINCIPLE. There is a close and very important interrelation between the Rutherford-Bohr atomic model and the periodic table of elements (Table 6·1).<sup>6·2</sup> The nucleus charge, which is equal to the unit charge  $e$  of the electron multiplied by the number  $Z$  of electrons surrounding the nucleus, known as the atomic number, increases in regular uniform steps on passing from one element to the next in order of atomic weight. Thus, hydrogen and helium with atomic numbers 1 and 2 consist of a nucleus of, respectively, one and two positive unit charges surrounded by the same number of electrons. As the atomic number increases, both the positive charge and the number of electrons increase.

It might be assumed that the normal state of an atom is such that all electrons would occupy positions of lowest energy, that is, orbits of lowest quantum number. However, this is not the case; the actual distribution of electrons over the stationary state orbits is governed by *Pauli's exclusion principle*.<sup>6·3</sup>

Considering the manner in which the electrons of the various elements occupy the possible stationary-state orbits of the model, it is evident that, in the case of the hydrogen atom, the single electron, following the principle that the stable state is one of lowest energy, will occupy the energy level defined by the quantum number  $n = 1$ . This is, however, not the only quantum number required to identify the state of the electron; having two degrees of freedom, the motion of the electron is determined by the energy and the angular momentum, both of which are quantized, that is, can only change discontinuously. Whereas the first or *principal quantum number n* determines the energy

level which is an integral multiple of  $\epsilon$ , the second, *angular-momentum* or *orbital quantum number*  $l$  is related to the angular momentum of the motion of the electron relative to the nucleus; it is an integral multiple of  $h/2\pi$  and expresses the "ellipticity"

TABLE 6·1  
PERIODIC SYSTEM OF ELEMENTS

								1	H			2	He
1. Period	3	4	5	6	7	8	9					10	Ne
	Li	Be	B	C	N	O	F						
2. Period	11	12	13	14	15	16	17					18	Ar
	Na	Mg	Al	Si	P	S	Cl						
3. Period	19	20	21	22	23	24	25	26	27	28			
	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni			
	29	30	31	32	33	34	35					36	
	Cu	Zn	Ga	Ge	As	Se	Br						Kr
4. Period	37	38	39	40	41	42	43	44	45	46			
	Rb	Sr	Y	Zr	C'b	Mo	Ma	Ru	Rh	Pd			
	47	48	49	50	51	52	53					54	
	Ag	C'd	In	Sn	Sb	Te	J						Xe
5. Period	55	56	57-71	72	73	74	75	76	77	78			
	Cs	Ba	R.E.	Hf	Ta	W	Re	Os	Ir	Pt			
	79	80	81	82	83	84						86	
	Au	Hg	Tl	Pb	Bi	Po							Rn
6. Period		88	89	90	91	92							
		Ra	Ac	Th	Pa	U							

of the orbit, that is, the ratio of the major to the minor axis. As the first quantum  $n$  varies from  $n = 1$  to  $n = \infty$ , the second quantum number  $l$  varies from  $l = 0$  (circular orbit) to  $l = n - 1$ . Thus in the simple Bohr model the lowest quantum state of the electron is described by the two quantum numbers:  $n = 1$ ,  $l = 0$ . The next higher orbit of quantum number  $n = 2$  (which

would be occupied by the single hydrogen electron only in an unstable *excited* state) is associated with two possible quantum states of angular momentum;  $l = 0$  and  $l = 1$ ; similarly for  $n = 3$ , the angular-momentum quantum number can take the three values:  $l = 2$ ,  $l = 1$ , and  $l = 0$ . The number of electrons necessary to complete an orbit of principal quantum number  $n$  consisting of  $(n - 1)$  subgroups defined by the quantum number  $l$  is thus  $2n^2$ .

It has been found that two additional quantum numbers are required to describe the orbital motion of atoms consisting of more than one electron since the simple Bohr model with two degrees of freedom of electron motion is no longer adequate for their representation. These two numbers are called the *magnetic quantum number* and the *spin* quantum number.<sup>6-4</sup> The former defines the deviation from the assumed elliptical orbital motion because of the precession of the elliptical orbits, resulting from the relativistic change of mass of the electrons with changing velocity. The latter defines the rotatory motion or *spin*, superimposed on the orbital motion, which is associated with the motion of every type of particle considered to have a finite extension. As the spin is described by the *direction* of the angular momentum, the spin quantum number can have only two values, one positive and one negative. The magnetic quantum number, on the other hand, can take all values between  $+l$  and  $-l$  including zero.

With the aid of all four quantum numbers and Pauli's exclusion principle the occupation of the stationary state orbits by any number of electrons can be explained. This principle simply stipulates that no two electrons belonging to any individual nucleus may have the same set of four quantum numbers. After one orbit has been occupied by the maximum number of electrons that can be accommodated according to the exclusion principle, additional electrons occupy orbits of greater radii and smaller binding energy. Thus the *impenetrability* of matter which is manifested by the small compressibility of solids is essentially a consequence of the exclusion principle.

Since both the magnetic moment and the spin have a very small influence on the energy level, and the spin quantum number can take only two values of opposite sign, a satisfactory approach to many problems concerning the quantum states of electrons

is possible by using only the first two quantum numbers  $n$  and  $l$ . In terms of these two quantum numbers the maximum number of electrons that may simultaneously occupy a suborbit of any given momentum quantum number  $l$  is given by the expression  $2(2l + 1)$  where  $(2l + 1)$  denotes the number of possible values of the magnetic (third) quantum number for each successive value of  $l$  and the factor 2 expresses the double value of the spin quantum number. Hence for  $n = 1$  and  $l = 0$ , the number of electrons occupying the lowest orbit is 2. The two electrons of

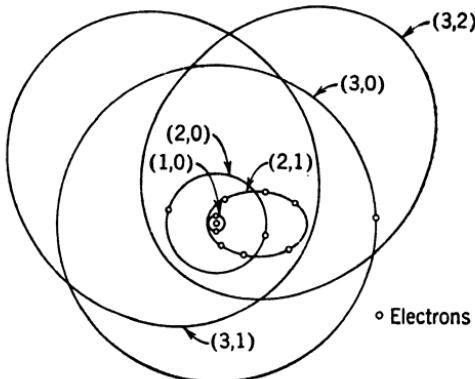


FIG. 6·1 Schematic representation of electron orbits of sodium atom, defined by quantum numbers  $(n, l)$ .

the helium atom will therefore both occupy the same orbit with  $n = 1$ ,  $l = 0$ . The third electron of lithium (atomic number 3) can, however, no longer be accommodated in this orbit and must enter the next higher orbit of principal quantum number  $n = 2$ , with which two different quantum numbers of angular momentum  $l = 0$  and  $l = 1$  are associated. Since the orbit defined by  $(n = 2, l = 0)$  can be occupied by 2 electrons, the orbit defined by  $(n = 2, l = 1)$  by  $2(2 + 1) = 6$  electrons, the energy level defined by the principal quantum number  $n = 2$  can be occupied by  $6 + 2 = 8 = 2n^2$  electrons. Hence, the atom of the element with  $2 + 8 = 10$  electrons which, according to the periodic table is neon, fills both energy levels  $n = 1$  and  $n = 2$  completely. Figure 6·1 is a schematic illustration of the orbits of sodium (atomic number 11) in two-dimensional projection.

The stable or *inert* character of both neon and helium, as well as of such other elements of higher atomic numbers whose elec-

trons completely occupy all orbits defined by a principal quantum number, such as argon with atomic number 18, krypton with atomic number 36, xenon with atomic number 54, all of which are called *inert gases*, is the result of the stability of the grouping of electrons in what are usually called *closed electron shells* formed by the completely occupied electron orbits of quantum number  $n$  in accordance with Pauli's exclusion principle.

The order of succession in which the electrons occupy the various orbits is determined by the energy expended in the process of binding successive electrons. Thus, if, after the completion of a secondary shell of angular momentum number  $l < (n - 1)$  within the total group of such shells associated with the principal quantum number  $n$ , the binding energy of an electron in the first  $l$  shell of the  $(n + 1)$  level is higher than that of the successive  $l$  shell in the  $n$  level, the first  $l$  shell of the  $(n + 1)$  level will be occupied by electrons before the still unoccupied  $l$  shells of the  $n$  level will be completed. This consideration explains the apparent irregularities in the filling of orbits appearing in Table 6·2, which indicates the configuration of the electron shells of atoms in their order in the periodic table.

Electrons that occupy orbits outside of closed secondary or  $l$  shells and whose number is always less than 8 which is the minimum number of electrons forming a closed secondary shell, are called *valence electrons*. Of all the electrons of the atom only the valence electrons determine the interaction between atoms, and thus the chemical and mechanical properties of the material. The similarity of chemical and mechanical properties of various elements is intimately related to the similarity of the structure of their outer electron shell, that is, to the identity of the number of their valence electrons. This is the explanation of the occurrence of periods in the system of elements. Thus, for instance, the alkali metals (lithium, sodium, potassium, rubidium, caesium) being all chemically and mechanically similar, (they are, for instance, more compressible than other solids) all have one electron more than the preceding inert gas, that is, have one electron outside of the last closed shell. Similarly, the halogens, (fluorine, chlorine, bromine, iodine) showing all nearly identical chemical reaction, have one electron less than the inert gas that follows them in the periodic table, that is, each has an outside shell to the completion of which one electron is lacking. The

inert gases, on the other hand, have all closed shells of 8 electrons; there is therefore no atomic interaction in inert gases.

According to Pauli's exclusion principle, individual stable groupings of electrons in orbits of nearly equal energy are made up of 2, 8, 18, and 32 electrons (see Table 6·2). These are

TABLE 6.2  
FILLING OF ELECTRONS SHELLS OF ELEMENTS

Period	Quantum Numbers	$n =$	1	2	3		4			5			6	7	
		$l =$	0	0	1		0	1	2	0	1	2	3	0	1
	1H - 2He	1-2													
1	3Li -10Ne	2	1-2	1-6											
2	11Na-18Ar	2	2	6	1-2	1-6									
	19K -20Ca	2	2	6	2	6		1-2							
3	21Sc -28Ni	2	2	6	2	6	1-8	1-2							
	29Cu -36Kr	2	2	6	2	6	10	1-2	1-6						
	37Rb -38Sr	2	2	6	2	6	10	2	6			1-2			
	39Y -40Zr	2	2	6	2	6	10	2	6	1-2		2			
4	41Cb -45Rh	2	2	6	2	6	10	2	6	4-8		1			
	46Pd	2	2	6	2	6	10	2	6	10					
	47Ag -54Xe	2	2	6	2	6	10	2	6	10		1-2	1-6		
	55Cs -56Ba	2	2	6	2	6	10	2	6	10		2	6		1-2
5	57-71(R.E)	2	2	6	2	6	10	2	6	10	0-14	2	6	1	2
	72Hf -78Pt	2	2	6	2	6	10	2	6	10	14	2	6	2-8	2
	79Au -86Rn	2	2	6	2	6	10	2	6	10	14	2	6	10	1-2
6	87 -92U	2	2	6	2	6	10	2	6	10	14	2	6	10	2
															1-6

therefore the numbers of electrons that make up the closed electron shells; they are very significant in the formation of the chemical and mechanical properties, as they determine the number of valence electrons.

The forces which two atoms exert on each other are the result of the electric charges of their valence electrons. The orbits of

electrons in *closed shells* remain practically unaltered by the influence of neighboring atoms. When two atoms approach each other closely enough, the valence electrons which are at the largest distance from their own nuclei take up new common orbits around the nuclei of both atoms, thus establishing an *atomic bond*. The strength of this bond is the expression of the mutual influence between the valence electrons and the nuclei of the bound atoms.

The incomplete outer shell of an atom can be closed by the addition of the required number of electrons; or the valence electrons outside the closed shell, which are rather loosely connected with the nucleus, can be removed from the atom by the supplying of a certain amount of energy. The value of this energy is an indication of the ease with which the atom of the considered element lends its valence electrons in chemical reactions. As a result of the loss of the valence electrons, the electrically neutral atom becomes positively charged; as a result of the completion of the outer shell by the occupation of the still available orbits by electrons, it becomes negatively charged. Being electrically charged, the atom becomes able to move or "wander" within an electric field. Charged atoms with closed electron shells are therefore called *ions* (Greek for "wanderer"); the positively charged ions are known as *cations*, the negatively charged as *anions*.

## 7. Quantum Statistics and "Wave Particles"

**THE WAVE FUNCTION.** The Bohr model in its classical form is inadequate to describe atoms more complex than the hydrogen atom, unless additional concepts are introduced, such as the spin and Pauli's exclusion principle. These two concepts are, however, less a "refinement" of the initial Bohr model than an expression of the fact that this model is intrinsically inadequate to describe the structure of a complex atom. Although originally introduced to support the Bohr theory, these concepts actually form part of the modern quantum-statistical theory of atomic structure, which has supplanted the Bohr model.

The inadequacy of Bohr's model is due to the fact that it represents a rather arbitrary fusion of classical and quantum-mechanical concepts. It treats the rotating electrons by methods of classical dynamics of moving bodies, into which the quantum

concept of selective orbits is injected axiomatically from the outside.

The new quantum-statistical theory of matter is based on the realization that

1. A simple extension of classical mechanics to the description of phenomena involving physical entities of very small order of magnitude is not possible, since only in classical mechanics can a particle be exactly defined by its position and velocity.

2. The condition of *quantization* of frequencies or states of energy must be contained in the concept of atomic structure and not arbitrarily imposed on it.

The first consideration is contained in the concepts of quantum statistics and in Heisenberg's "uncertainty principle"; the condition of automatic quantization is fulfilled by the de Broglie-Schrödinger-Dirac "wave-particle" theory.<sup>7·1</sup>

In the wave-particle theory every attempt at a pictorial representation of the atom has been abandoned. The representation of an elementary particle as a minute mass or a *store of energy*, defined by its position in space and by its velocity, has been replaced by an unpicturable embodiment of particle and wave characteristics, by the aid of which the dual particle-wave character of matter can be described and the particle and wave theories unified. Because of this *dual* character of matter, light behaves in certain respects as if it were made up of electrically neutral moving, electron-like light particles, called *photons*, whereas in other respects it behaves as if it were a wave. On the other hand, particles of matter, such as electrons and nuclei, have a wave nature as well as that of a particle, since they show the same phenomena of diffraction and interference that are characteristic of light waves. This fact has been proved in various ways, the most important of which is the electron microscope in which the bombardment of a crystal surface with electrons produces the wave-diffraction pattern characteristic of X rays and light beams.<sup>7·2</sup> Thus, observation of radiation phenomena tends to support the assumption of the particle character and observation of interference phenomena that of the wave character of light as well as of matter.

Light and matter are linked together by the fact that they are both forms of energy; the observed dual character of light has a parallel in the same dual character of matter. These observa-

tions provide the basis for the picture of a "particle" which, pulsating rhythmically, spreads this pulsation in the form of a wave. These *wave particles* have a wave length  $\lambda$  associated with them, which is given by the *de Broglie equation*  $\lambda = h/mv$ , where  $mv$  is the momentum of the particle of mass  $m$  and relative velocity  $v$ . Thus, a beam of light is regarded as a stream of moving photons; its intensity is expressed by the density of these photons, which can be described in terms of a wavelike function. A similar function describes the density of an electron shower by which a metal surface is bombarded in the electron microscope. It is in terms of such *wave functions* that the diffraction phenomena associated both with the photons and with the electrons can be described, although no perceptible vibrating medium is involved. From the consideration that every *particle* is accompanied by a *wave* without attempting to form a definite picture of either the particle or the wave or even of the nature of the medium in which this wave is propagated, it follows that the motion, that is, the displacement  $y$  of whatever is vibrating is governed by a *wave equation*, the solution of which gives the wave function.

The simplest one-dimensional form of a wave equation is the well-known expression governing oscillations of a single degree of freedom  $y'' + c^2y = 0$  where  $c$  denotes a constant. The solution of this equation is the wave function,

$$y = A \sin cx + B \cos cx \quad (7 \cdot 1)$$

Every solution of this form satisfies the wave equation; it does not, however, describe the behavior of any particular system, as such a system must have definite boundaries. If, as in the vibration of a string of length  $s$ , the conditions  $y = 0$  for  $x = 0$  and  $x = s$  are introduced as the boundary conditions of the vibrating system, the values of the constants become  $B = 0$  and  $A \neq 0$  if the trivial solution  $y = 0$  is to be avoided. The constant  $A$ , however, can take only a number of definite discrete values; because of the imposed second boundary condition  $y = A \sin (cs)$  must be zero for  $x = s$  or, since  $A \neq 0$ ,  $\sin (cs) = 0$ . Hence  $cs = n\pi$  or  $c = (n\pi/s)$ . The possible waves are therefore expressed by the equation  $y = A \sin n\pi$ , where  $n$  can take all integral values. This last equation is the condition of *quantization* of frequencies; it follows directly from the wave character

of the phenomenon and the necessity of fulfilling given boundary conditions.

The quantum condition of the Bohr model can also be easily interpreted in terms of the wave picture. Considering the circular motion of an electron around a nucleus, the quantum condition of angular momentum is given by the expression  $mvr = nh/2\pi$ . If instead of the revolving electron a revolving wave is considered, the previous quantum condition of the particle is

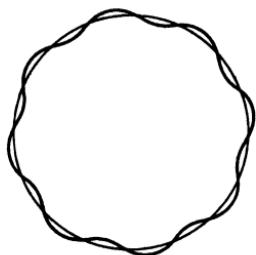


FIG. 7·1 Wave motion around a circle.

transformed into the quantum condition of the wave by introduction of the de Broglie equation  $mv = h/\lambda$ . Hence,  $n\lambda = 2\pi r$ , which is the expression of the requirement that the stable motion of a wave along a circular orbit is only possible when the circumference of the circle is an integral multiple of the wave length (Fig. 7·1). Thus, the quantization follows automatically from the wave character of the electron.

Schrödinger's *wave equations*<sup>7·3</sup> of the electron or rather of the vibrating "essence" are of a similar form with energy as the constant. Their solutions will therefore be those that correspond to discrete values of the energy for which the boundary conditions of the considered space are fulfilled; the possible energy levels of the electron are thus quantized. When Schrödinger's wave equation is applied to the hydrogen atom, the quantized energies of the Schrödinger waves correspond to the discrete energy levels (orbits) of the Bohr model. Higher energy levels correspond with waves of shorter lengths. The frequencies at which energy is emitted or absorbed are the *beat* or *interference frequencies* of consecutive wave functions.

**THE STATISTICAL CONCEPT.** So far no attempt has been made to attach any physical meaning to the vibrating "essence" which is associated with the electron and is described by the wave equation. The interpretation of the wave function  $\psi$  requires the introduction of statistical concepts expressing the limitations in the validity of the concept of causality.

The validity of this concept, that is, the existence of causal relations is implied in the use, in Newtonian mechanics, of differential equations, by which the unknown relations between certain

quantities at a certain moment are derived from the known relations at a preceding moment. As long as the principle of causality is valid, the final state of a system must follow unambiguously from its original conditions. In spite of the general validity of the laws of Newtonian mechanics, their practical applicability is restricted to the analysis of phenomena for which the momentary conditions as well as all acting influences not only are known in principle, but also can at any moment be actually described. Therefore, if the investigated phenomenon is the resultant effect of a large number of phenomena, which, although individually governed by Newtonian mechanics, cannot be described in terms of Newtonian mechanics, the causal approach must be abandoned. In phenomena of this type which are called *mass phenomena*, the influences governing individual phenomena are so varied and complex that any attempt to study each phenomenon separately would be impracticable; practical methods of approach to the analysis of the resultant phenomenon can therefore only be statistical.

A good illustration of a *mass phenomenon* and of the implications of the statistical approach in mechanics is provided by the *Galton board*. This is a vertical board, studded with a large number of parallel rows of metal pins which are staggered by one half of the pin distance. Small steel balls, released from a certain height on the center pin of the top row, will, on hitting this pin, be deflected to either the right or the left. Each ball, after being so deflected proceeds vertically on its downward way until it hits a pin of the second row and must again choose between right and left, and so on. When the ball has thus traversed all  $x$  rows, it has made this choice  $x$  times and as a result finds itself at a certain horizontal distance to the right or left of the center. This procedure is repeated either with a large number of similar balls or a large number of times with the same ball, the number of balls at a certain distance from the center or the various positions of the single ball being recorded.

The assumption that the actual distribution of balls or the various positions of one ball at the bottom of the Galton board can be explained in a causal way, because the motion of each individual ball is governed by Newtonian mechanics, is an illusion. In attempting to follow every motion of an individual ball, considering all effects during the fall and in the collisions

with the pins, where extremely small force components decide about the future course of the ball to the right or left of the pin, the individual phenomenon will become too complex to be analyzed. However, the probability that the ball will be diverted to the left or to the right by the individual pin that it strikes can be assessed beforehand by purely logical argument to be  $1/2$ , if the conditions (distances of pins, size of balls, and the like) are considered to be perfectly uniform. From this assumption the distribution of a large number of balls at the bottom of the board, or the probability of a certain position of an individual ball, can be derived by the theory of probability.<sup>7-4</sup>

With increasing number of balls their distribution at the bottom will be bell-shaped, tending towards the distribution described by the normal or Gaussian curve. This curve represents the distribution function for a practically infinite number of balls; the ordinates express the actual number or the percentage of balls that, passing through all the pin rows, aggregate at a certain number of unit distances from the center. This percentage also represents the theoretical probability of finding an individual ball at the respective distance.

It is only in the form of probabilities and distribution functions that results of statistical theories can be represented. These functions express the probability of finding any one of the large number of elements involved in the considered process or phenomenon in a certain state, or the number or percentage of elements that can be expected to be in this state. Statistical theories are unable to predict the exact behavior of a particular single element and are meaningless in this respect. Information with respect to an individual element is thus represented either by the *a priori* probability of a definite behavior, obtained by the consideration of the list of the various possible behaviors, or by the probability of a certain behavior derived from a previously obtained experimental distribution function involving a large number of observations or elements.

Suppose the picture of the ball falling on the Galton board is replaced by the picture of a particle moving in space and characterized at any moment by its position and its momentum (which is proportional to its velocity). Heisenberg has pointed out that it is impossible to observe or measure simultaneously and with the same accuracy both the position and the velocity

of such a particle, since the act of observation itself interferes with one of the coordinates. Light of a wave length sufficiently short to determine accurately the position of a particle of electron size has also enough energy to *deflect* the electron from its course during the observation, when it is hit by photons. It is thus given an additional momentum in an unspecified direction, in the same way in which the impact of the *observation* of the position of the falling ball on the Galton board, that is, the hitting of the pin, produces the individually unpredictable deviation from its path. Increasing the wave length of the light, that is, decreasing the momentum of the photons, in order to produce less interference with the electron under observation, necessarily reduces the accuracy in the observation of its position. Thus, the conditions that favor the accurate measurement of the position produce the greatest uncertainty in the velocity, and vice versa.

Both the impact of the light radiation on the electron and of the collision with the pin on the falling ball are purely statistical phenomena. The "chance" effects of these impacts on the motion, therefore, cannot be eliminated by additional safeguards or additional accuracy. They illustrate the indeterminate nature of all problems of observation and of measurements in microphysics, which is expressed by Heisenberg's *uncertainty principle*. This principle also follows directly from the wave-particle concept of matter. If, under certain conditions, the wave particle is observed as a wave, it can have more than one possible location. On the other hand, if it is observed as a particle in a definite location, its wave length and thus its velocity, which is related to the wave length by the de Broglie equation, cannot be measured.

The uncertainty relation is not confined to location and velocity; it is also valid in the coordination of energy and time. Conditions that produce the greatest accuracy in the measurement of energy are those that produce the greatest uncertainty in time. Thus the measurement of the energy of a particle at a certain instant is associated with an uncertainty which is the larger the shorter the time of measurement.

As a result of Heisenberg's uncertainty principle no particle of atomic or subatomic size can be visualized, since it cannot have a definite position and a definite velocity at the same time.

These variables can only be expressed in terms of the probability of finding the particle within a certain range of locations or velocities.

Because of the large number of electrons involved, their location can only be defined by the probability of finding a certain percentage of them in a definite location. It is this large number of electrons, the positions of which are given in terms of a probability function, that produce the wave pattern of behavior, described by Schrödinger's equation. Thus, the *wave function*  $\psi$  is a *probability function*, and the waves are the *waves of the probability* of finding a particle in a definite position or at a certain energy level.

In the case of light the interference between different wave patterns produces different light intensities, which are measured by the square of the amplitude of the resultant wave. Similarly, in the case of probability waves, the *intensity* of the particle distribution, that is, the relative density of particles in a certain location, which is proportional to the probability of a particle being in this location, is measured by the square of the wave function. Thus the probability-distribution function for the position of the electron, or the *electron density* in a certain volume  $dV$ , is obtained as the product  $\psi^2 dV$ . It is a measure of the probability of finding the electron in the volume  $dV$ , or of the volume density of electrons or of the distribution of their charge.

For the Bohr model of the hydrogen atom the probability of finding the electron between the distances  $r$  and  $(r + dr)$  from the nucleus is expressed by  $4\pi r^2 \psi^2 dr$ . This function has a maximum at  $r = r_0$  which is identical with the Bohr radius of the orbit  $n = 1$ . However, according to the quantum-statistical concept the electron is not restricted to this distance alone but can take other though less probable positions.

Suppose the probability of finding the electron at a certain distance  $r$  from the nucleus or the specific density of electrons or the electron charge at this distance is considered as a cloud; the density of this cloud varies. The *cloud density* for the hydrogen atom would increase from the nucleus to a maximum beyond which it would thin out rather rapidly, although even at a large distance a very slight "probability haze" would still exist. For atoms of more than one electron, each *cloud*, taken as a whole, represents a single, fully occupied energy level of electrons,

which forms a *closed shell*. These *clouds* are not actually separated from each other, but rather form one cloud of fluctuating density, as illustrated in Fig. 7·2, which shows the distribution of the electron density or electron charge in the hydrogen atom and in the two shells of a sodium atom, as a function of the distance from the nucleus.

The Bohr model of atomic structure can thus be reinterpreted in the light of quantum statistics by considering that the electrons move in and out about the nucleus, remaining usually

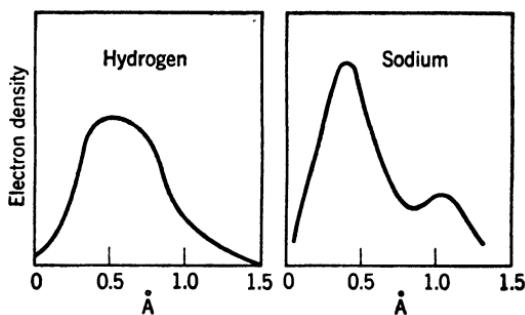


FIG. 7·2 Distribution of electron density around nucleus of hydrogen and of sodium atom.

within a distance that does not differ appreciably from the Bohr orbit; however, the probability that electrons may be found at appreciably different distances is finite, no matter how small. The velocity of the motion is not constant; however, it is of the order of magnitude of the constant Bohr velocity. Over a period of time, long enough to permit a large number of cycles of the motion of the electrons around the nucleus to take place, the atom can thus be visualized as consisting of the nucleus, surrounded by a spatial *electron cloud*, which actually represents the positions of the individual electrons, that is the electrons blurred by the "time exposure" of their rapid motion in the period of observation. The density of this "photograph" of the electron cloud will be highest at the location at which the probability of finding an electron is largest.

The retention of the essential Bohr picture is very helpful in visualizing the atomic structure of matter for which this model has so far provided the best illustration. The *statistical* picture of the atom is, however, no longer the planetary system with

fixed orbits around the nucleus, but a nucleus surrounded by cloudlike shells of electrons or, rather, of electron densities. Although the density of this (statistical) cloud does not abruptly drop to zero beyond a certain radius but decreases gradually towards the exterior, this decrease is rapid enough to justify the representation of atoms and of ions as nuclei surrounded by electronic clouds, which, in very rough approximation, can be assumed to be of spherical shape. Thus the external radius of the cloud represents the ionic or the atomic radius.

This picture of the atom shows a striking resemblance to the old picture of the atoms considered as elastic spheres out of which material is formed in the same manner in which such spheres are *packed* together. Although the "substance" of the sphere has greatly changed, and its size, electric charge, and manner of interaction with other spheres can now be derived from quantum statistics, the simple outward picture has remained almost unchanged.

## 8. The Size of Atoms and Molecules

The size of particles forming the material is a most important characteristic, since both the intensity of thermal motion and the order of magnitude of the interacting forces are associated with particle size which, in turn, determines the particle distances in the group formation. The whole range of particle sizes, the ranges of applications of observation methods, and the ranges of different types of motion are represented in Fig. 8·1.

Comparison of the order of magnitude of the atomic nucleus ( $10^{-12}$  micron) and of electrons ( $10^{-10}$  micron) with the diameter of the hydrogen atom ( $10^{-4}$  micron) confirms the dimensional adequacy of the "solar-system" atomic model. The diameter of the orbit of the motion of the electron about the nucleus is about  $10^6$  times its own size. The radii of ions are derived from the interatomic spacing and represent the radii of the outer closed electron shells (or rather clouds); they are of an order of magnitude of 0.5 to 2.5 Å. If considered in the order of their atomic numbers, these radii show a periodic variation, in conformity with the periodicity of their electron structure as expressed by the periodic table; the largest volume in each sequence is occupied by the inert gas. This fact can be explained by a consideration of the relative charges of the nucleus and of

the electrons of the respective ions. If for instance the third sequence (quantum number  $n = 3$ ) is considered, starting with argon (atomic number 18) the electrons of which are distributed over three shells, it is clear that in the ion of potassium (atomic number 19), which is obtained by the removal of 1 valence electron and has therefore only 18 electrons and a positive charge,

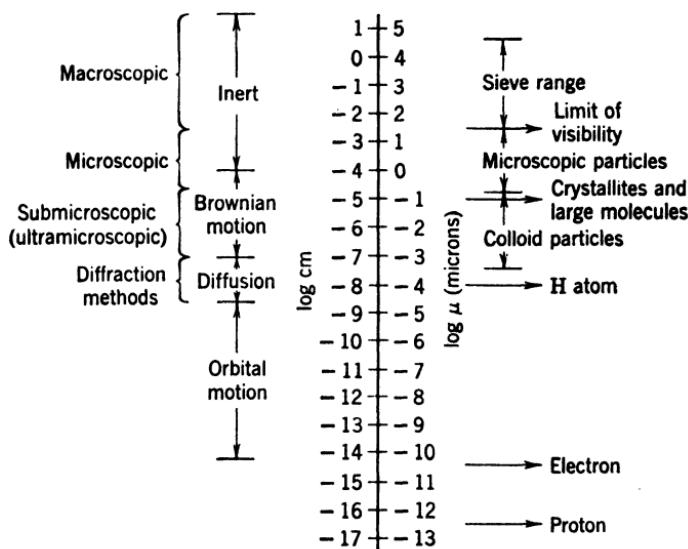


FIG. 8·1 Sizes of material particles, types of motion, and methods of observation.

the nucleus has one positive excess unit charge. This produces a stronger bond between nucleus and electrons than exists in the neutral argon atom, in which the electric charges are balanced. The increased bond manifests itself by a decrease of the volume occupied by the electrons. For the same reason the ion of calcium, with atomic number 20, 18 electrons, a positive charge, and a difference of two units between the charge of the nucleus and of the electrons, will occupy a still smaller volume, and so on. When the shell is again closed in the formation of an inert gas, the difference between the charges of nucleus and electrons becoming zero, the volume of the ion which is identical in this case with that of the atom, suddenly expands.

The radii of the atoms are larger than the radii of the respective

ions by roughly the thickness of the outer electron shell in which the valence electrons are located (Fig. 8-2). The varying tendency of the relation between atomic radius and atomic number within the periodic sequences, which differs from the definitely falling trend in the relation between the ionic radius and the atomic number, is due to the charges of the ion. The single-valence electron of potassium, which makes the radius

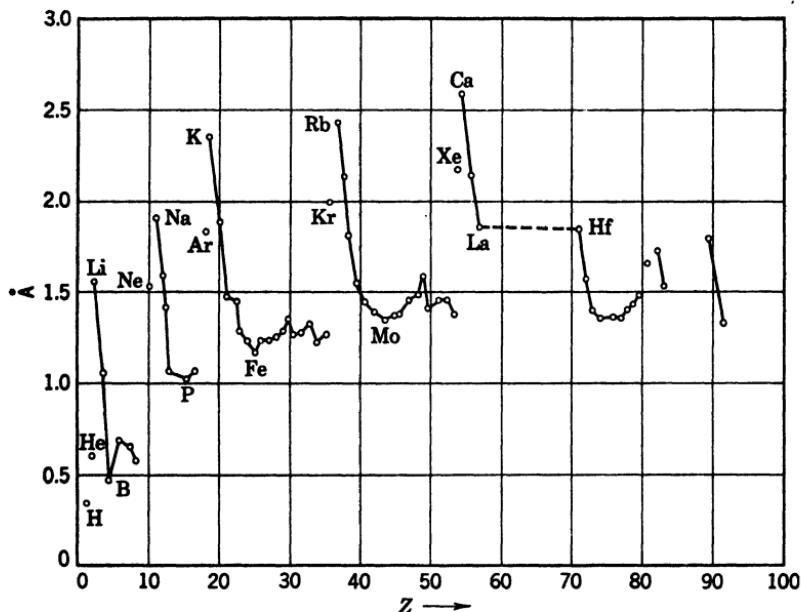
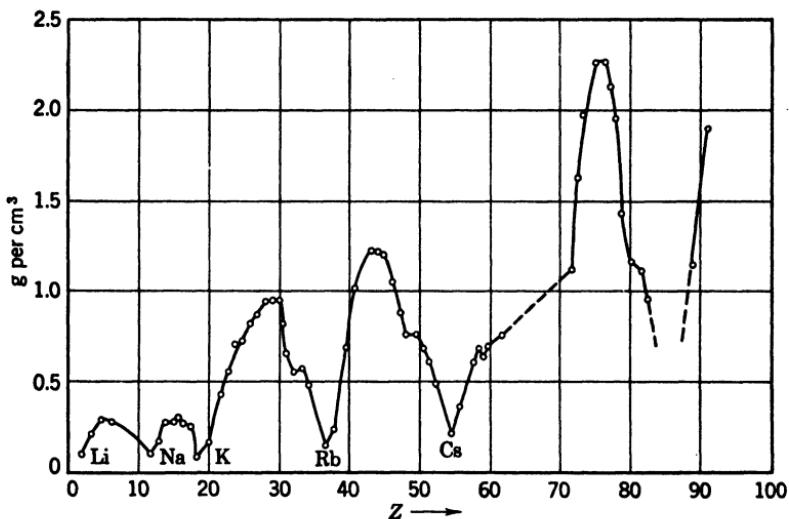


FIG. 8·2 Radii of atoms as a function of atomic number  $Z$  (after Zwikker<sup>8·1</sup>).

of the potassium atom larger than that of argon, is bound by one positive unit charge to the ion; the two valence electrons of calcium, while occupying the same orbit are, however, attracted by two unit charges; they will be more closely bound and therefore will occupy a smaller volume. Thus, within the first part of every sequence the atomic radii decrease, in spite of increasing number of electrons; the atomic volume reaches a minimum for elements in the middle of the periodic sequences. In the second half of this sequence, however, the repulsive forces between the swarming electrons apparently counteract the tendency towards decreasing volumes.

Figure 8·3 represents the relation between the specific density and the atomic number of elements and is easy to interpret on the basis of Fig. 8·2. It illustrates the close relation between atomic radius and specific density in solids and tends to justify the rather simple picture of the solid, consisting of a more or less closely packed aggregation of atoms or of ions of spherical shape; this is true at least for solids in which the structure is neither the



According to this picture of the crystal structure the particles with the smallest radii would be the easiest to arrange between more or less closely packed atoms of larger radii. Since it can generally be observed that decreasing interatomic distances are accompanied by increasing strength of interatomic bonds, these small particles would form the strongest bonds but would at the same time be the most mobile within the packing of atomic spheres of larger diameter. Ions or atoms of such small size could thus escape or *diffuse* out of the closely packed larger atoms, without disturbing the stability of their arrangement. Figure 8·2 shows that the smallest particles are those of hydrogen boron, carbon, nitrogen, and oxygen. These are, therefore, the elements easiest to fit into or to diffuse out of a stable arrangement of other atoms and most likely to form the strongest bonds. Their atomic radius is of the order of magnitude of  $0.5 \text{ \AA}$ , whereas the atomic distances of most of the metals are between 1.5 and  $3.0 \text{ \AA}$ .

If spheres of uniform radius  $r$  are packed very loosely so that each layer of spheres lies immediately above the one below in cubical arrangement in which the spheres touch each other in 6 points, the maximum radius  $r_1$  of the spheres that can be introduced into the voids between those spheres is  $r_1 = (\sqrt{3} - 1)r = 0.73r$ . In the most densely packed aggregate of spherical particles of cubical type, in which the interstices between the spheres are bounded by six spherical surfaces and in which the spheres touch each other in 12 points, smaller particles of a radius  $r_1 = (\sqrt{2} - 1)r = 0.41r$  can be introduced into the interstices without disturbing the original arrangement of spheres.<sup>8·2</sup> In experimental studies of a number of alloys it was found however<sup>8·3</sup> that a limit of  $r_1/r = 0.59$  could be reached; apparently the lattice is capable of a certain adjustment by expansion or by a slight deformation of the electron shells, the "impenetrability" of which is not absolute. The relatively high radius ratio  $r_1/r$  of the closely packed structures explains the relatively high mobility of the small atoms of carbon, nitrogen, and oxygen within most of the crystalline materials, particularly metals, as observed in diffusion processes. Thus the interstices of closely packed iron atoms with an interatomic distance of  $2r = 2.50 \text{ \AA}$  can be filled by spheres of a radius  $r_1 = 0.41 \times 1.25 = 0.5 \text{ \AA}$ . The radii of carbon and nitrogen are larger than this figure,

whereas the radius of oxygen atoms is slightly smaller; thus a certain expansion will accompany the insertion of carbon or nitrogen into the iron lattice, whereas oxygen can diffuse in and out of the lattice without affecting it.

For radius ratios exceeding 0.59, simply packed structures no longer exist. The existing structures are more complex and, in fact, form transitions from the structures in which the alloying atoms are absorbed into the interstices to those in which they are substituted for particles of the original lattice. Iron-carbon alloys form actually borderline cases since the radius of the carbon atom slightly exceeds the limiting value of  $r_1 = 0.59 \times 1.25\text{\AA} = 0.74 \text{ \AA}$ . This is the reason for the complex transformations taking place at the "austenitizing" temperature of 720°C below which a perfect substitution alloy ("austenite") does not exist. These transformations are responsible for the hardening of steel.

It should be noted that the atomic radius of a particular element in different crystals is not constant but depends on the type of atomic forces present. Thus, for instance, the radius of an ion of the same metal will be smaller in an ionic crystal than in a metal crystal. Hence, in specifying the atomic radius the type of crystal from which it has been obtained must be known, and a comparison of atomic radii is limited to crystals with the same type of binding forces. There is, however, also a dependence on the crystal lattice such that the atomic radius increases with increasing coordination number.

The differences of atomic radii of different metals are relatively small. Also where a metal can exist in various lattices (allotropy), the difference in spacing between the different modifications is small. Thus, the different modifications have all about the same energy, which is the explanation of the fact that allotropy is very common among metals.

The relative magnitude of atomic radii is thus a very important effect in the formation of metal alloys, since, in order to form alloys, metal atoms either must be of such relative dimensions as to produce a structure in which the interstices of the packed larger atoms are filled by smaller atoms (*interstitial alloys*) or must be nearly the same size (*substitutional alloys*). It has been found that in the latter case the radii must not differ by more than 15 percent,<sup>8-4</sup> the smaller this difference, the easier the formation

of the alloy. No substitutional alloys are formed of elements if the diameter of atoms differs by more than 15 percent.

If the electron clouds of two or more atoms lose their character as isolating shells and, penetrating each other, form a stable group in which the charges of the valence electrons are neutral-

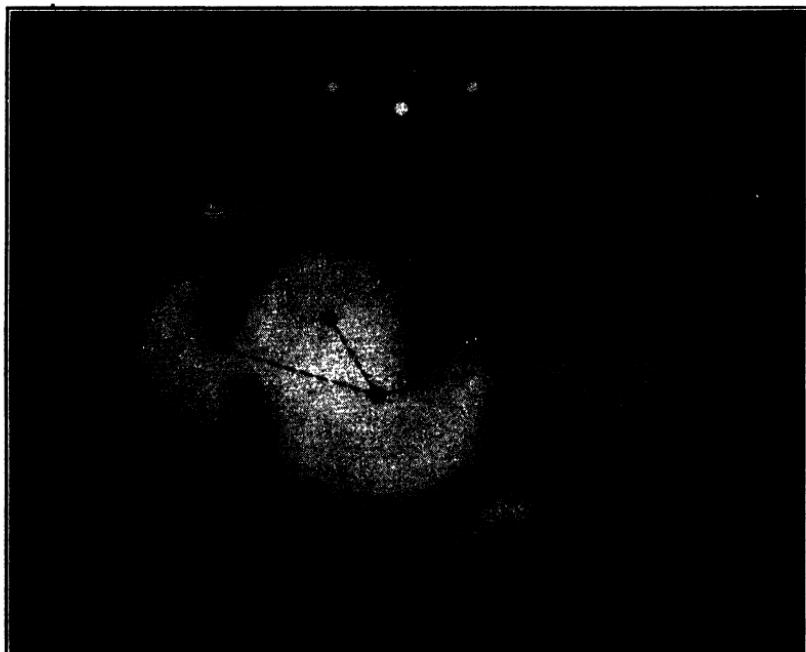


FIG. 8·4 Quantum-statistical model of four-atomic ammonia molecule  $\text{NH}_3$  (courtesy American Institute of Physics<sup>8·5</sup>).

ized in binding the atoms together, such a group is called a molecule (Fig. 8·4). Because of the neutralization of the charge of the individual group of atoms, the forces of interaction between the molecules will be relatively small, whereas the atomic bond forces forming the molecule will be large. Although molecules are thus electrically inert, they form a sort of temporary or permanent magnet called a *dipole* as a result of the fact that, within the formed molecule, the centers of gravity of the positive and the negative charges of the ions do not coincide. Such oscillating dipoles, whose magnetic or *dipole moment* is equal to the pole charge times the distance between the centers of the charge, attract each other and form intermolecular bonds.

In defining the molecule in the solid state as a specific chemical entity, the structural aspect of the group formation is neglected. In many cases the formation of a specific *chemical* entity is not associated with the formation of a specific *structural* entity and thus is irrelevant in those considerations where only the fact of the existence of a *structural* entity is of consequence. Thus, the formation of *chemical molecules* in all crystalline substances is usually irrelevant in relation to their mechanical behavior. The crystal lattice is formed by the ions occupying the lattice points, not by the chemical molecules theoretically formed by these ions. These chemical molecules are interconnected within the crystalline structure by strong interatomic bonds and have therefore no structural identity. In the sodium chloride lattice, for instance (Fig. 8·5), atoms of sodium and of chlorine are assumed to form molecules of sodium chloride which are considered chemically identifiable; such identification is, however, structurally meaningless, since the structure consists of the undivided crystal lattice. Thus, molecules should be considered as separate "particles" in a structural sense only if they form definite structural groups in such a manner that the mechanical separation of the individual groups (molecules) from each other is considerably easier, than the separation of individual elements (atoms) from the groups. According to this definition, only the "organic" or carbon compounds, and materials in which atoms aggregate into small groups before building up the structure, such as a number of silicon compounds and sulphur, form real structural molecules in solids.

Only in the liquid state has the chemically defined molecule also a structural identity, since in solutions the atoms have no independent stable existence outside of a structurally identifiable molecule.

The magnitude of molecules varies widely with the number of

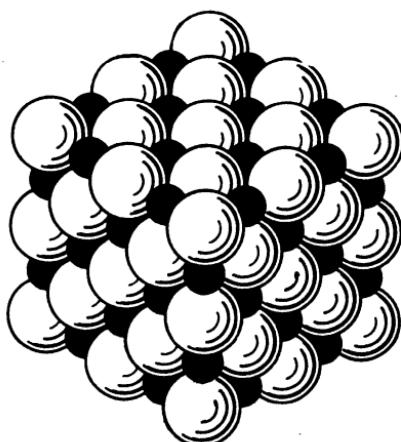


FIG. 8·5 Sodium chloride lattice.

atoms forming them. It is therefore meaningless to define the intermolecular distances in terms of particle size, since these distances will be characteristic of the individual molecular structure rather than of the molecule itself. The dimensions of single molecules are usually of the order of magnitude of 1.5 to 5 Å. However, the so-called giant molecules or macromolecules forming the high polymers and fibers attain considerably larger dimensions. The number of atoms forming a molecule varies between 2 and  $10^3$  for single molecules of an order of magnitude  $<10^{-7}$  cm, to  $10^9$  for macromolecules whose order of magnitude reaches  $10^{-5}$  cm.

The mechanical properties of materials built up of structurally identifiable molecules are affected not only by the size, but also by the particular shape of the molecules. This influence of the shape of the constituent molecules on the phenomenological behavior increases with increasing size of the molecules; it is relatively small if the shape of the molecules does not deviate too much from the sphere, but becomes predominant where fiber-like macromolecules are formed by interconnected oriented molecules of elongated shape.

### References

- 5·1 A. L. PIPES, *Applied Mathematics for Engineers and Physicists* McGraw-Hill Book Co., New York (1946) 172.
- 5·2 *Ibid.*, 177.
- 5·3 M. BORN, *Atomic Physics*, Blackie & Son, London (1945) 77.
- 5·4 F. O RICE and E. TELLER, *The Structure of Matter*, John Wiley & Sons (1949) 5.
- 6·1 M. BORN, *op. cit.*, 98.
- 6·2 *Ibid.*, 35.
- 6·3 *Ibid.*, 168.
- 6·4 *Ibid.*, 145.
- 7·1 *Ibid.*, 84–97.
- 7·2 F. O. Rice and E. TELLER, *op. cit.*, 41.
- 7·3 M. BORN, *op. cit.*, 130.
- 7·4 R. v. MISES, *Wahrscheinlichkeitsrechnung*, F. Deuticke, Leipzig (1930) 144.
- 8·1 C. ZWIKKER, *Technische Physik d. Werkstoffe*, J. Springer, Berlin (1942), 4, 5.
- 8·2 D. R. HUDSON, *J. Applied Phys.* **20** (1949) 154.
- 8·3 G. HAEGG, *Z. physik. Chem.* **12** (1930–31) 33.
- 8·4 R. H. HARRINGTON, *The Modern Metallurgy of Alloys*, John Wiley & Sons, New York (1948) 138.
- 8·5 *J. Applied Phys.* **20** (1949), cover plate February issue.

## THE STRUCTURE OF MATTER

**9. Forces of Interaction**

**TYPES OF BONDS.** The energy of cohesion of a solid is represented by the energy absorbed in forming the solid out of atoms or molecules and is equivalent to the "mass defect." It is therefore equal to the *heat of sublimation*, that is the heat energy required to dissociate the unit quantity of the solid into isolated free atoms or molecules. The forces of interaction between the particles form the interatomic and intermolecular bonds in which the cohesive energy of the material is contained.

The unit quantity of a substance is usually defined either as a *mole* or as a *gram-atom*, that is the quantity the weight in grams of which is equal to the molecular or to the atomic weight of the substance; these are, respectively, the weights of a molecule or of an atom of the substance. These weights are, however, not measured in grams but with reference to a standard gas, the atomic weight of which is designated as unity and which is defined in such a manner that the atomic weight of oxygen is 16. A mole of diatomic oxygen molecules thus weighs 32 grams. From the definition of the mole and the gram-atom it follows that the unit quantity always contains, respectively, the same number of molecules or of atoms. This number is a physical constant known as *Avogadro's number*, and has the value  $N = 6.02 \times 10^{23}$ .

It is convenient to consider four principal types of bonds, although such classification is not rigorous. In fact the transition from one type of bond to another is so gradual that bonds

of intermediate types, combining the characteristics of some of the principal types, must necessarily exist. The four principal types of bonds are:

1. *Ionic* or electrostatic bonds, also called heteropolar bonds.
2. *Atomic* bonds, also called valence, covalent, or homopolar bonds.
3. *Metallic* bonds.
4. *Intermolecular* bonds, formed by dipole or Van der Waals' forces.

The intermolecular bonds should not be grouped together with the three other bond types, since by acting between individual molecules they are acting at a different level of association of particles. To distinguish them from the other three types, known as the *primary* bonds between atoms or ions, intermolecular bonds are usually referred to as *secondary* bonds. It has already been pointed out that, as a result of the neutralization of the electric charges of the ions in the formation of molecules, and because of the comparatively large intermolecular distances, the intermolecular forces have a considerably lower energy content than the interatomic or interionic forces and are therefore much weaker.

The energy content of bonds is generally an inverse, though not linear, function of the distance between the particles. The difference in energy content between primary and secondary bonds is approximately of two orders of magnitude, whereas the difference between interatomic and intermolecular distances is not more than one order of magnitude.

Not all bonds between molecules are intermolecular or *secondary* bonds. The molecular segments out of which the long chain or macromolecules of high polymers are formed are connected by bonds formed between adjoining atoms of two molecules and are therefore primary bonds. Only the bonds between molecular chains themselves are secondary, with the result that the transverse strength of high polymers of fibrous structure made up of parallel macromolecules is only a fraction of its strength in the direction of the molecular chains.

All four types of bonds may be responsible for the cohesion of crystalline solids, although two basically different types of formation can be distinguished. In the first type, structurally identi-

fiable molecules are formed by primary bonds, and grouped into a lattice in which they are held together by secondary bonds; these are the so-called *molecular crystals* or *molecular lattices*. In the second type of solid no molecules are formed, or the molecules are only chemically, not structurally, identifiable; their separation from each other thus requires the application of energy of the same order of magnitude as the separation of the bonds between the individual atoms. Formations of this type are called *coordination lattices*.

Either amorphous substances are formed by primary bonds alone, the chemical molecules being structurally unidentifiable, as in most of the strong glasses; or chemically and structurally identifiable molecules are held together by secondary bonds, as in most of the organic compounds, such as high polymers.

The study of interacting forces between particles in liquids has not yet led to any well-established concept. It is, however, assumed that forces of a similar type are responsible for the formation of a material both in the solid and in the liquid states. Thus, the melt of materials of ionic-lattice structure is probably formed by the same electrostatic (primary) forces of interaction that are responsible for the solid lattice formation. The same is apparently true for atomic lattices, although the operative mechanism of the interatomic forces in the liquid is not known. There are, however, some cases where the transition from the solid to the liquid state is accompanied by a definite change of the type of operating bonds, as in a number of complex chemical compounds. Most liquids, however, retain their molecular structure after solidification.

**IONIC BONDS.** Ionic bonds result from the electrostatic interaction (Coulomb attraction) between (positive) cations and (negative) anions. They produce the *ionic crystals* or anorganic salts, the formation of which is based on a one-to-one relation between heteropolar neighboring particles. Since the atoms of metallic elements lose their valence electrons very easily, at least as long as the number of electrons outside the closed shell is very small, whereas atoms of nonmetallic elements with very nearly closed outer shells have a strong tendency to complete this shell by absorbing electrons, stable cations and anions are formed by the association of metal and nonmetal atoms. These ions of opposite charge are mutually attracted and form ionic

bonds, which are of considerable strength. The energy content or heat of sublimation of ionic crystals is of the order of magnitude of 100 to 500 kg cal per mole. Ionic-bond formation is the easier, the smaller the number of valence electrons of the constituent atoms; it is therefore particularly frequent between the monovalent alkali metals and the halogens, which lack one electron in their outer shell; sodium chloride is a typical example of an ionic crystal. Evidently, ionic bonds are formed only between atoms of different chemical elements.

In order for atoms of the same elements to be combined, they must have incomplete shells; otherwise, only molecular bonds are formed, as illustrated by the behavior of the inert gases. Between atoms of the same element both *covalent* and *metallic* bonds may be formed. The nature of the bond that is actually formed depends on the number of valence electrons.

**COVALENT BONDS.** Covalent bonds form the *valence crystals*, which are monatomic substances of high cohesive energy, great hardness, and very low conductivity. Diamond, carborundum, and quartz are prototypes of such crystals. The heat of sublimation of the ideal valence crystals is of an order of magnitude similar to that of ionic crystals. Atomic or covalent bonds are formed by the "sharing" of valence electrons by neighboring atoms which attempt by this process to complete their outer electron shells, thereby becoming negatively charged. According-



FIG. 9·1 Structure of material formed of diatomic molecules.

ing to Pauli's exclusion principle, a stable electron shell, beyond that associated with the quantum number  $n = 1$  (hydrogen and helium) contains at least 8 electrons. If there are, for instance, 7 electrons in the outer shell, as in the halogens (fluorine, chlorine, bromine, iodine), one electron is *shared* or *exchanged* by both atoms. These atoms have therefore alternately 6 and 8 electrons in their outer shells, becoming in

alternation positively and negatively charged. Through the *sharing* process which produces an exchange of electric charge, the atoms form a closely knit two-atom or *diatomic* molecule. Because the mutual interaction within the diatomic molecules extends to pairs of atoms only (Fig. 9·1), the cohesive strength

of such materials is small; in crystal formation they are held together by weak molecular bonds.

In elements with 6 electrons in the outer shell such as sulphur, selenium, or tellurium, one electron of each atom is shared with the two nearest neighboring atoms, so that a "roving pair" of electrons is exchanged among three atoms forming a molecule. Since each atom can thus form a bond with only two neighboring atoms, a linear chain of atoms can be formed (Fig. 9-2), producing molecules that may become very large. Such one-dimensional fiber-, ring-, or spiral-shaped giant molecules resulting in a "threadlike" structure are characteristic of crystalline sulphur, selenium, and tellurium. As with diatomic molecules, only molecular interaction is possible between the linear molecules; the cohesive strength of materials in this formation is therefore relatively small, unless the molecules are oriented in the direction of the force.

If there are 5 electrons in the outer shell, as in arsenic, antimony, and bismuth atoms, the stable 8-electron shell of an atom

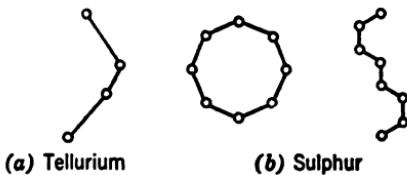


FIG. 9·2 Linear (spiral or ring) molecules.

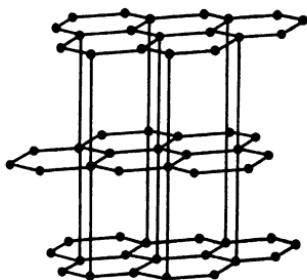


FIG. 9-3 Nearly plane giant molecules of antimony or graphite of hexagonal structure.

is completed by the sharing of one electron each with the three closest neighbors, so that each atom in the four-atom group is alternately surrounded by 8 electrons. In this way, plane or wavelike giant "sheet" molecules are formed which consist of layers of atoms which are usually connected by molecular bonds (Fig. 9.3). The molecular interaction produces a rather weak cohesion of the material in the direction perpendicular to the planes of the molecules, resulting in a two-di-

dimensional "sheetlike" structure, strong only in the planes of the sheets.

Atoms of elements with 4 electrons in the outer shell, such as

carbon, silicon, and germanium share one electron each with the four closest neighbors, producing a three-dimensional formation in which the atoms occupy the center and the corners of a tetrahedron (Fig. 9·4). Because in this formation, unlike the other formations discussed so far, the forces of interaction in all three directions of space are interatomic forces, the rigidity and strength of such substances are very high, as exemplified by the diamond

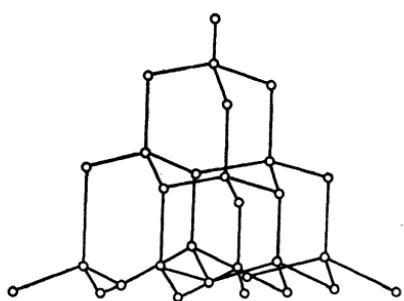


FIG. 9·4 Three-dimensional diamond structure.

and silicon crystals. Moreover, the fact that the 4 shared electrons, by alternately closing the outer shells of the constituent atoms, produce configurations characteristic of the electrically inert gases (carbon atoms with atomic number 6, and 4 shared electrons are transformed into the 10-electron neon-like configuration; silicon atoms with

atomic number 14 are transformed into the 18-electron argon-like configuration and germanium with atomic number 32 into the 36-electron krypton-like configuration) is responsible for the poor electric conductivity of the materials which form the so-called diamond-like atomic structures.

In general, if  $N$  denotes the actual number of valence electrons in the incomplete shell,  $(8 - N)$  will be the number of electrons to be *shared* by neighboring atoms, so as to complete the shell alternatively in any one of the neighboring atoms, connected by covalent bonds. Thus, the number of *unshared* electrons is  $N - (8 - N) = (2N - 8)$ ; this number can evidently not be less than zero, as otherwise each atom would have to share more electrons than it possesses. Hence,  $N = 4$  is the minimum number of valence electrons leading to covalent bonds; this means that only materials consisting of atoms having at least 4 electrons in the outer incomplete shell can form covalent bonds.

The electrons that are jointly held or shared by the bound atoms are assumed to be effective in alternately completing the closed shell for each atom by some mechanism of "multiple-duty performance" or of "mutual interchange." Since ions formed by the momentary completion of the outer electron shells

are negatively charged, atom groups formed by covalent bonds are homopolar.

It has been shown that the cohesive strength and the deformational behavior of materials formed by covalent bonds are essentially determined by the number of bonds one atom can form with its neighbors. The larger this number, the higher becomes the total cohesive energy of the material, and the nearer does the atomic structure approach the rigid three-dimensional diamond-like configuration.

Covalent bonds also form the molecules of the organic compounds. These structurally identifiable molecules are very tightly knitted and form the structural units of an immense number of organic compounds including the high polymers. They are built up of the three *electronegative* (that is, electron-hungry) elements: carbon, nitrogen, and oxygen, with a ballast of hydrogen. Covalent bonds are also responsible for the building of molecules into long parallel or intertwined chains or fibers by a process of spontaneous intermolecular combination, called *polymerization*, as well as for combining them into three-dimensional strongly interlinked skeleton structures, such as the *cross-linked* polymers and the silicates (glasses).

The nature of the covalent bond has been pictured as the taking up by the *shared* valence electrons of orbits common to all atoms participating in the bond. This picture does not consider the quantum-statistical concepts of the structure of matter. In terms of quantum statistics *sharing* of electrons could be expressed by specifying either definite probabilities of finding the *shared* electrons in the space around the bound ions or a definite density distribution of a valence-electron cloud surrounding the participating ions. Considering a covalent bond between two atoms, that is, a single pair of valence electrons shared by two atoms, the density distribution of the valence-electron cloud surrounding the bound atoms is obtained by coordinating the electron clouds of individual atoms in such a way that the shared electrons can take positions around either of the nuclei. Since the density of any electron cloud is given by a wave function associated with a certain energy level, it is necessary to find a combined wave function compatible with the normal states of both atoms which corresponds to a minimum value of the total energy, that is, to an energy value lower than the energy of the wave functions of

the isolated atoms and also lower than the energy of any alternative wave function. This combined function is in fact a coordination of the two wave functions into a *resonant* system.<sup>9-1</sup>

Thus the establishment of a covalent bond is associated with a *resonance* between the probability or electron-density waves around both ions, by which the valence-electron cloud is made to move periodically between the ions. This is a phenomenon somewhat like the interchange of vibration energy within the system of two resonating oscillators discussed in Art. 5. The energy exchange between the oscillators is analogous to the exchange in charge associated with the interchange of electrons between the atoms. The picture of the tuning forks referred to in Art. 5 gives an approximate idea of the quantum-statistical concept of resonance and of the nature of the covalent bond between two atoms.

If more than two atoms are connected, *resonance* is established between the wave functions of all atoms participating in the bond.

**METALLIC BONDS.** If the number of valence electrons of an atom is not large enough ( $N < 4$ ) to form covalent bonds, and if the electrostatic charges of the ions are not of opposite sign, so that the formation of ionic bonds is excluded, the only remaining possibility of interaction is by a common "pool" of the valence electrons. The ions have given up their valence electrons and have thereby acquired a positive excess charge, whereas the electrons have lost their connection with individual ions and have become *free*. Hence, according to the *free-electron* theory, a metal is composed of a regular array of positive ions which are immersed in a "gas" of free electrons (Fig. 9·5). These electrons, being no longer confined to the individual ions, move freely between them. Since in this formation the ions have no direct link with one another, metal atoms are nondiscriminating and therefore very versatile in the formation of bonds. Any single atom will therefore act on as many neighboring atoms as can be crowded into the space around it. Thus metal atoms will always attempt to have as many neighbors as possible, a tendency that must lead to densely packed structures. This conclusion is borne out by the fact that the characteristic metal structures are formed by the two possible types of *closest packing* of spherical ions.

Usually metals are divided with regard to their chemical composition into the monatomic metals and the alloys. Another significant division is with respect to the filling of secondary electron shells of orbital quantum number  $l = 2$  into *simple metals* with completely filled or completely vacant shells, and into *transition metals* with partly filled shells (see Table 6·2). The structure and properties of the transition metals are usually more complex than those of the simple metals; moreover, their

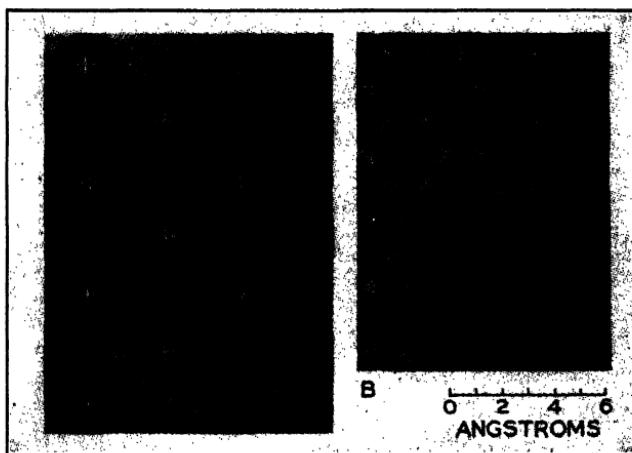


FIG. 9·5 Metallic structure of (A) sodium and (B) copper (after Shockley<sup>9·2</sup>).

cohesive energy (heat of sublimation) is markedly higher. The heat of sublimation of monatomic simple metals is of the order of magnitude of 10 to 150 kg cal per mole; that of the monatomic transition metals varies between 70 and 220 kg cal per mole. The fact that the  $l = 2$  shell below the level of the valence electrons is incompletely filled is responsible for many of the characteristic properties of the transition metals. Thus, for instance, the ferromagnetism of iron, cobalt and nickel, which are transition metals (see Table 6·2), is attributed to the existence of the incompletely filled  $l = 2$  shell in the  $n = 3$  level; it has also been found that other elements containing unfilled  $l = 2$  shells are either ferromagnetic or strongly paramagnetic. In elements in which these levels are filled, as in copper, which follows nickel in the periodic table, ferromagnetic properties do not exist.

Most of the monatomic metals crystallize in the simple densely packed structures. Complex structures that involve mainly atoms of higher valence are very similar to those of the valence crystals, discussed previously. Thus, there is a gradual transition from the covalent to the metallic binding. The most characteristically metallic elements have, in general, not more than 2 valence electrons. With increasing number of valence electrons the truly metallic binding is gradually replaced by covalent binding. In certain elements, therefore, both types of bond exist simultaneously. Examples are lead with 4 valence electrons, which nevertheless forms a densely packed cubic structure of predominantly metallic type, and tin, also with 4 valence electrons, which, at low temperature, forms a nonmetallic diamond-type (covalent) structure ("grey tin"); at room temperature, an essentially metallic structure of relatively dense packing ("white tin"). In such cases it is assumed that only part of the bound atoms is fully *ionized*, that is, has given up all its valence electrons to the free electron fog, while others have retained part of their valence electrons, which form covalent bonds.

The versatility in establishing metallic bonds and the lack of the restrictions with regard to the number of bonds and direction of bond formation between individual atoms, characteristic of the ionic and covalent bonds, produce the typical metal properties of ductility and malleability. In ionic and covalent formation irrecoverable changes in the relative location of ions interfere permanently with the established atomic bonds; however, bonds in metals are easily reformed, as no particular relative position of atoms is favored in the bond formation, and no stabilization of bonds by the completion of electron shells or by the sharing of electrons has taken place.

A model of the crystal structure of a metal has been developed by Bragg,<sup>9-3</sup> in which the metal is represented by an assembly of soap bubbles, approximately one millimeter in diameter, floating on the surface of a soap solution. The model represents the behavior of a metal very closely, because the bubbles are uniform in size, are held together by the surface tension of the soap solution which represents the binding forces of the free electrons, and the bubbles glide past each other practically without friction if a shearing force is applied. Most of the effects characterizing the metal structure can be simulated

with the aid of this model, such as grain boundaries, dislocations, and lattice defects, slip, crystal fragmentation, and the influence of foreign atoms. This can be seen from Fig. 9·6 in which a photograph of an ideally regular assembly of bubbles is reproduced; Fig. 9·7 is a model of the real arrangement of atoms in the vicinity of boundaries between individual crystals.

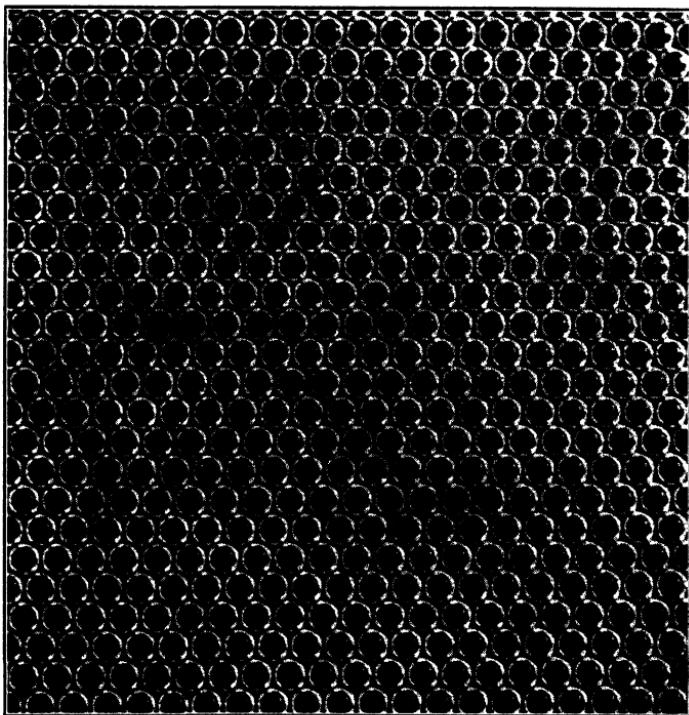


FIG. 9·6 Soap-bubble model of regular metallic structure (courtesy Sir Lawrence Bragg<sup>9·3</sup>).

**COMPLEX BONDS.** So far it has been assumed that only one type of bond exists in any material. Actually several types of bonds may exist simultaneously, for example covalent and ionic, or metallic bonds: some of the valence electrons may form covalent bonds, whereas the rest forms ionic or metallic bonds. The properties of such materials are very much affected by this mixed type of bond formation. Graphite, mica, talcum, asbestos, and textile fibers are examples of materials formed by mixed bonds. Thus, while a diamond which is chemically identical

with graphite is formed by all 4 valence electrons entering into covalent bonds, only 3 valence electrons are so connected in graphite, whereas the fourth forms a metallic bond between the trivalent carbon ions having a positive unit charge. The interatomic distances associated with the covalent bonds is about  $1.42 \text{ \AA}$ . As a result of the particular type of bond, the molecule formation of graphite should be similar to that of a chemical element with 5 valence electrons, such as antimony. The simi-

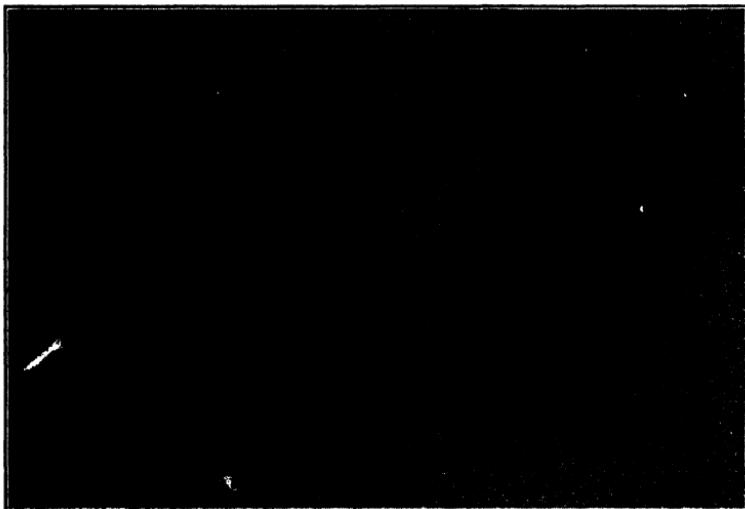


FIG. 9·7 Soap-bubble model of metallic structure including boundaries between crystals (courtesy Sir Lawrence Bragg<sup>9·3</sup>).

larity is immediately evident if the graphite lattice is compared with the antimony lattice (Fig. 9·3). The covalent bonds produce the strength of the sheetlike molecules of the graphite which is almost the same as that of diamond; the remaining free electrons are responsible for easy deformability and low strength between the molecular layers spaced at about  $3.4 \text{ \AA}$ , which makes the use of graphite as a lubricant possible. Mica is formed of sheetlike molecules built by covalent bonds from silicon and oxygen atoms, which are held together by ionic bonds, whereas the sheetlike talcum molecules formed by covalent bonds from the same atoms are held together by intermolecular bonds producing the particularly large spacing of  $18.8 \text{ \AA}$ .

Even simple ionic bonds such as those forming brittle sodium

chloride crystals may be transformed into a complex structure of ionic and covalent or metallic bonds if, for instance, the sodium atoms are replaced by silver atoms with partially filled  $l = 2$  shells in the formation of ductile silver chloride.

**INTERMOLECULAR BONDS.** Intermolecular (secondary) bonds, in general, are produced by the magnetic interaction between dipoles, which may be rapidly changing and due to the motion of the electrons, or permanent and characteristic of the individual molecule, or temporarily induced in electrically neutral molecules by an external electric field surrounding an ion or another dipole. Intermolecular bonds produce loosely bound lattices of saturated molecules. In the solidified inert gases the bound lattice elements are atoms, but the interacting forces are of the order of magnitude of the intermolecular forces, as no other bond is possible.

Intermolecular bonds are responsible for the formation of molecular crystals, such as paraffin, or crystallized rubber, and of the organic compounds, such as polymers, or gels. The heat of sublimation of the truly molecular compounds is of the order of magnitude of a few kilogram calories per mole. In those substances the shape of the molecules, which are usually macromolecules, is the principal factor influencing their mechanical behavior and properties.

**INTERATOMIC AND INTERMOLECULAR FORCES.** Independently of the character of the possible bonds between particles, as presented in the foregoing discussion, a general schematic approach may be applied to the treatment of the interatomic or intermolecular forces by considering any type of bond to be produced by the interaction of two potential fields of repulsive and attractive energy, or of two fields of repulsive and attractive electrostatic charges. In ionic bonds these attractive forces are the result of the opposite charges; in covalent bonds they may be considered the result of *sharing* of valence electrons; in metallic bonds they may be considered produced by the general attraction of the positive ions for the negatively charged "free-electron fog" which holds the ions together. The repulsive forces resisting the too close approach of ions may be considered as arising from the *impenetrability* and the resulting small compressibility of the closed electron shells or from the electrostatic repulsion between charges of the same sign. An additional source of repulsion is

the thermal oscillation of particles around their equilibrium position. At room temperature the amplitude of the thermal oscillations is of an order of magnitude of less than  $\frac{1}{10}$  of the atomic distance; it increases, however, appreciably (two- to threefold) at elevated temperatures.

Under the assumption that the variation of the potentials of the attractive and repulsive forces with the distance from a particle may be expressed in the general form  $(\text{const}/r^n)$ , the resulting potential energy of interaction  $\phi(r)$  has the form,<sup>9·4</sup>

$$\phi(r) = -\frac{a}{r^m} + \frac{b}{r^n} \quad (9\cdot1)$$

$n > m$  since the repulsive potential  $(b/r^n)$  must decrease more rapidly with increasing distance  $r$  than the attractive potential  $(a/r^m)$ , as otherwise no equilibrium would be possible. The negative sign of the attractive energy potential is due, as explained in Art. 6, to the fact that the value of the energy potential must be negative for  $r > 0$ .

The actual values of  $m$  and  $n$  vary with the nature of the bond. For electrostatic attraction the potential  $(-a/r^m)$  is equal to  $(-e^2 v_1 v_2 / r)$ , where  $v_1$  and  $v_2$ , respectively, denote the number of valence electrons of charge  $e$  in the two bound atoms. Thus  $m = 1$  is generally introduced for primary bonds. The value of  $n$  varies between 9 and 11 for ionic and covalent bonds, and between 6 and 9 for metallic bonds, although for alkali metals this value may be as low as 3. For secondary bonds the exponents usually introduced are  $m = 6$  and  $n$  between 9 and 12, or even larger.<sup>9·1</sup>

If eq. 9·1 represents the combined potentials of the attractive and repulsive forces, these interacting forces are given by the expressions:

$$F(r) = \frac{d}{dr} \phi(r) = \frac{a}{r^{(m+1)}} - \frac{b}{r^{(n+1)}} \quad (9\cdot2)$$

The equilibrium position of the two interacting particles is attained at the distance of maximum stability, that is, of minimum (negative) energy at the bottom of the "potential trough" drawn schematically in Fig. 9·8. At the point where  $\phi(r)$  reaches a minimum, its first differential  $F(r) = 0$ , indicating equilibrium between the interacting forces (Fig. 9·9).

The equilibrium position is defined by the minimum (negative) potential energy or maximum bond energy  $u_0$ ; the change of potential with the displacement of one particle from that position is schematically illustrated by the diagram of the energy potential  $\phi(r)$ . The energy required to remove the particle out of the field of mutual interaction is given by the area of the function  $F(r)$ , which thus represents the bond energy.

From a consideration of this function it is evident that any displacement from the equilibrium position gives rise to restoring forces which, for small displacements, may be assumed to increase linearly with the displacement. The linearity fails, though, as soon as the displacements are appreciable. The restoring forces differ for both directions of the displacement. Figure 9.8 shows that the restoration or elastic constant, which is determined by either the slope of the tangent to the  $F(r)$  curve, or the curvature of the  $\phi(r)$  curve at the equilibrium position  $r = r_0$ , in-

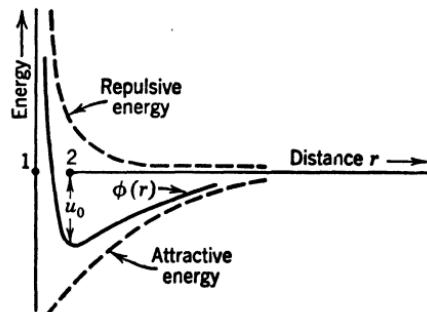


FIG. 9.8 Diagram of bond energy  $\phi(r)$  (after Houwink<sup>9-5</sup>).

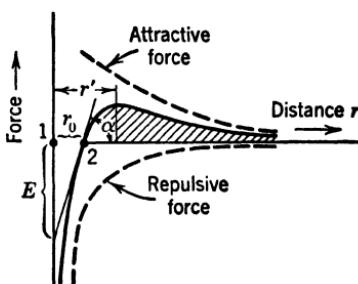


FIG. 9.9 Diagram of interatomic force  $F(r)$  (after Houwink<sup>9-5</sup>).

creases rapidly if the particles are crowded together but decreases gradually with increasing separation. The maximum force required to separate the particles is given by the maximum ordinate of the  $F(r)$  curve. This ordinate represents the theoretical or atomic cohesive strength of an elastic substance of perfectly regular atomic arrangement.

Although for small displacements the deviation from linearity of the relation between displacement and restoring force is small, the asymmetry of the  $F(r)$  curve with regard to its zero ordinate is sufficient to produce the thermal expansion of materials by

producing, for the same restoring force, a larger amplitude of the outward vibration than of the inward vibration, thus shifting outward the center of vibration, that is, the equilibrium position of the particles. Since the difference between the amplitudes producing the outward shift increases with increasing amplitudes, that is, with increasing temperature, the coefficient of thermal expansion necessarily increases with temperature, a conclusion that is borne out by observation.<sup>9-6</sup>

## 10. Thermal Oscillations

In groups built up of interacting atoms or of interacting molecules the individual atom or molecule behaves like a Planck oscillator. The oscillations that are important with regard to the mechanical behavior of materials, usually called the *thermal oscillations*, are the oscillations of the particles bound by the forces of interaction. The oscillations of the atoms within crystalline regions are also designated as *lattice vibrations*. The oscillations of molecules which are due to the intermolecular forces and to the oscillations of the atoms in the molecule are made up of translational and rotational components. The latter are the more important, the larger the molecule. The general interrelation between the coupling forces and the mode and frequency of oscillations has been illustrated by the two-oscillator model (Art. 5). In this model two modes of oscillation have been found to exist: the symmetric mode in which the oscillators move in the same direction, and the asymmetric mode in which they move against each other. These two modes represent limiting cases of the lattice vibrations obtained under the assumption that the length of the standing wave, in which the lattice vibrates, is infinite.

The vibrations of a lattice built up of a large number of oscillators, however, cannot be adequately discussed in terms of the two-oscillator model. It is the interrelation between the wavelength of the vibration of the lattice and the angular frequency of the particles oscillating around their positions of equilibrium that determines the character of the lattice vibration and the frequencies of radiation emitted by the lattice.

The characteristic behavior of the three-dimensional lattice can be illustrated by the analysis of the model of a one-dimensional lattice.<sup>10-1</sup> This model consists of a regular array of mass

points uniformly spaced along a line at distances  $d$  (Fig. 10·1), oscillating about their equilibrium positions with the angular frequency  $\omega = 2\pi\nu$ . The mass points are acted on by forces between the centers of particles, which tend to restore the equilibrium positions whenever particles have been displaced; it is assumed that only neighboring particles interact.

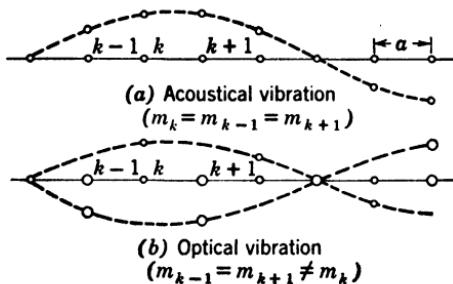


FIG. 10·1 One-dimensional models of vibrating crystal lattices.

For small displacements the force of restoration acting on any particle may be assumed proportional to the relative displacement of particles, thus replacing the curved shape of the  $F(r)$  function by its tangent at  $r = r_0$ , as shown in Fig. 9·9. Hence, the restoring force acting on any particle  $k$  depends on the momentary difference between the amplitude  $y_k$  of the particle and the amplitudes  $y_{k+1}$  and  $y_{k-1}$  of the neighboring particles.

The relation between the natural frequencies of the lattice vibrations and their wave length is obtained by solving the simultaneous equations of motion of all particles  $k$  of mass  $m_k$  subject to a restoring force  $F_k$  which is proportional to the difference of amplitudes. Hence,

$$F_k = \alpha(y_{k+1} - y_k + y_{k-1} - y_k) \quad (10·1)$$

and

$$m_k \frac{d^2 y_k}{dt^2} = \alpha(y_{k+1} + y_{k-1} - 2y_k) \quad (10·2)$$

which is identical with the differential equation of the vibrating string loaded with point masses  $m_k$ .

Two different types of eq. 10·2 can be set up by assuming either that the masses  $m_k$  are all equal or that the lattice consists of alternating light and heavy particles. With the first assumption neighboring particles will, at an average, move in the same

direction, whereas with the second assumption two types of motion are possible: one which is of the same type as before, and the second in which the motion of neighboring particles is, at an average, in opposite directions. This can be shown by solving the set of differential eqs. 10·2 for  $m_{k-1} = m_k = m_{k+1}$  and for  $m_{k-1} = m_{k+1} \neq m_k$ . In the first case one set of equations exist for all points  $m_k$  whereas in the second case two different sets exist for  $m_{k+1}$  and for  $m_k$ .

In order to obtain the velocity of propagation  $c$  along the linear array of masses  $m_k = m_{k+1}$ , of any disturbance of angular frequency  $\omega$  a general solution of eq. 10·2 is introduced in the form,

$$y_k = y_0 e^{i(\omega t - (2\pi kd/\lambda))} \quad (10 \cdot 3)$$

where  $\lambda$  denotes the length of the propagating wave. By introducing eq. 10·3 into 10·2 the following relation is obtained between the frequency  $\omega$  and the wave length  $\lambda$ :

$$m\omega^2 = 4\alpha \sin^2 \frac{\pi d}{\lambda} \quad (10 \cdot 4)$$

Because of  $\omega = 2\pi\nu = 2\pi \frac{c}{\lambda}$ , eq. 10·4 can be written

$$c^2 = \frac{\alpha}{m} \left( \frac{\lambda}{\pi} \right)^2 \sin^2 \frac{\pi d}{\lambda} \quad (10 \cdot 5)$$

For values of  $\lambda$  which are large in comparison with  $d$  the sine can be replaced by the argument; hence,

$$\omega = \sqrt{\frac{\alpha}{m} \cdot \frac{2\pi d}{\lambda}} \quad (10 \cdot 6)$$

and

$$c = d \sqrt{\frac{\alpha}{m}} = c_0 \quad (10 \cdot 7)$$

$c_0$  denotes the velocity of propagation for infinite wave length  $\lambda = \infty$ . The relation between  $\omega$  and  $\lambda$  and between  $c$  and  $\lambda$  for finite wave length, according to eq. 10·4, can be written in the form,

$$\omega = \frac{2c_0}{d} \left| \sin \frac{\pi d}{\lambda} \right| \quad (10 \cdot 8)$$

or, because of eq. 10·6,

$$c = c_0 \sqrt{\frac{\sin \frac{\pi d}{\lambda}}{\frac{\pi d}{\lambda}}} \quad (10 \cdot 9)$$

The angular frequency  $\omega$  thus varies with  $1/\lambda$  as indicated in Fig. 10·2, with a maximum.

$$\omega_{\max} = \frac{2c_0}{d} = 2 \sqrt{\frac{\alpha}{m}} \quad (10 \cdot 10)$$

for the wave length  $\lambda = 2d$ , which propagates at a velocity

$$c = \frac{2}{\pi} c_0.$$

In the material body the maximum velocity of propagation of a mechanical disturbance is the velocity of sound in the material

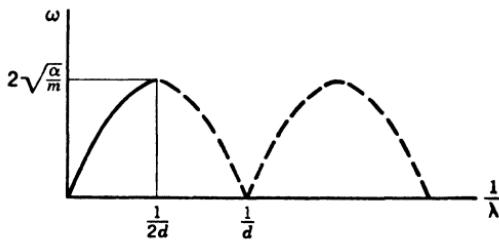


FIG. 10·2 Relation between angular frequency and wave length of vibration of linear lattice.

which, different for longitudinal and transversal vibrations, varies for metals between 300 and 600 meters per second. By introducing this value into eq. 10·10 and assuming that a value of  $d = 3 \text{ \AA}$  represents the average atomic diameter, a frequency

of the order of magnitude of  $\omega_{\max} = 2 \frac{300 \times 10^{10}}{3}$  to  $2 \frac{600 \times 10^{10}}{3}$

$= 2$  to  $4 \times 10^{12}$  is obtained as the order of magnitude of the frequency of lattice vibrations.

Oscillators vibrating at this frequency send out electrodynamic radiation of a wave length equal to the period of oscillation  $1/\omega$  times the velocity of light, which is  $3 \times 10^{10}$  cm per sec. The wave length of the radiation emitted by the vibrating lattice

is therefore of the order of magnitude  $10^{-2}$  cm or 100 microns. This is the *infrared* range, which extends from a wave length of a few microns to about 200 microns.

The very short infrared waves of length up to about 15 microns result from oscillations of the atoms within small molecules; they are not characteristic of the crystal lattice. The very long waves near the extreme infrared range of about 200 microns are essentially the result of rotational molecular oscillations. The intermediate wave length of 50 to 150 microns is associated with lattice oscillations as well as with oscillations of atoms around the mass center of complex large molecules.<sup>10-2</sup>

The frequencies of the infrared radiation are strongly affected by applied heat which changes both the frequencies of the radiation and its amplitude. It has been observed that the effect of applied heat is considerably stronger within the range of the relatively long waves than within the range of the short waves.

When, with increasing heat application the amplitudes of the lattice oscillations increase so much and so rapidly that they attain the order of magnitude of the interatomic or the intermolecular distances, that is, their kinetic energy attains the level indicated by the depth  $u_0$  of the *potential trough* (Fig. 9·8) the energy barriers between adjacent equilibrium positions can be momentarily overcome by heat vibrations; the particles thus activated by heat energy can either leave their respective equilibrium positions and occupy vacant positions of lower energy, or transmit energy by impact with and displacement of neighborhood particles. In the latter case the initial order of particles breaks down rapidly and the material *melts*. It is the essential difference between the solid and the liquid state that excessive thermal oscillations have destroyed the order in the lattice, facilitating place changes of particles. Although the interacting forces still tend to maintain a certain pattern of order, this tendency can not assert itself over any but the shortest range. This short-range order distinguishes the fluid from the gaseous state in which no tendency to create an ordered state exists because of the complete absence of attractive forces.

Melting starts at the lower temperature, the smaller the heat energy required to excite the oscillations to the amplitude at which place changes and displacements of particles by impact with other particles become possible. The start of melting

will also be the more abrupt, the more uniform the size and the more perfect the initial arrangement of particles in the crystal lattice. The less uniform the particles and the larger the differences between particle distances within the initial configuration of particles, the greater the possibilities of purely local breakdowns starting at points of least stability and spreading gradually over the whole structure, thus producing a very gradual start of the melting process. These considerations explain the lower and less well-defined melting points of amorphous materials in comparison with crystalline materials, as well as of materials of truly molecular structure in comparison with materials of atomic or ionic structure. The gradual change during melting of all mechanical properties within a relatively wide range of temperatures is a characteristic phenomenon of the so-called amorphous solids, by which they can be distinguished from crystalline solids, in which such changes are discontinuous.

The molecules oscillate and rotate around their positions of equilibrium in which they are held by the relatively weak intermolecular forces. Because of their comparatively larger masses and because of the considerably weaker forces of interaction, they oscillate at lower frequencies than atoms; these oscillations are also less regular. Their amplitudes are much larger than those of atomic oscillations because of the larger intermolecular distances. Because of the weak forces of interaction, which are easily overcome by heat vibrations, intensification of the thermal oscillations leads more easily to disintegration of the ordered arrangement in the molecular than in the atomic lattice.

Although the mode of oscillation of small compact nearly round molecules resembles to a certain extent that of the thermal oscillation of atoms, the modes of oscillation of long macromolecules will necessarily be different in the longitudinal and in the transversal direction. Because of the comparative ease of the transverse motion of the fiber-like molecules, such oscillations have the character of a wavelike wriggling motion rather than of an oscillation and rotation of a compact mass around a center. This mode of vibration governs the formation of chains of organic molecules by *polymerization*, which is the effect of chance contacts between vibrating molecules, as a result of which covalent bonds between atoms belonging to different molecules are established. As polymerization proceeds, the frequency of such

contacts decreases, because less reactive material, consisting of activated molecules is available. Polymerization thus proceeds at a decreasing rate; theoretically, it should not cease before all reactive molecules have been combined into chain molecules; however the increasing mechanical interference with the thermal motion of the individual molecules by already formed long chains stops the polymerization process long before the theoretical limit has been reached. Thus materials are obtained which consist of macromolecules in different *stages of polymerization*, that is, of different length and thermal activation.

For materials built up of atoms or small molecules the changes of properties in the melting process are reversible on solidification; melting of materials built up of chain molecules is associated however with a disruption of the long chains into shorter ones, which may produce irreversible changes of properties, though dissolution into individual molecules will only take place at temperatures exceeding the melting temperature.

The potential energy of the particles within any group formation is associated with the interacting forces in the position of equilibrium; the thermal or kinetic energy is the energy of the oscillations around this equilibrium position. From considerations concerning the possible degrees of freedom of the oscillating atoms and from the so-called *equipartition theorem* of statistical mechanics<sup>10-3</sup> (according to which each degree of freedom of the oscillating particle is associated with a mean potential energy of  $\frac{1}{2}kT$  and with the same amount of kinetic energy) the total energy content of a particle in space possessing three degrees of freedom is  $u = 3kT$ , where  $k = 1.39 \times 10^{-16}$  erg per °C is the Boltzmann constant, and  $T$  denotes the absolute (Kelvin) temperature. The total energy content per mole is  $U = 3RT$  where  $R = 1.96$  cal per °C is the "gas constant." The increase of energy of the particles per degree centigrade which is the heat required to increase the temperature of the material by one degree, represents the *specific heat* of a solid; it is therefore  $3 \times 1.96$  or approximately 6 cal per mole.

In solids the total energy content is made up of potential and of kinetic energy; the energy of a monatomic gas, consisting only of the kinetic energy of the three degrees of freedom of translatory motion is only one-half that of a solid. The energy difference of  $1.5RT$  is expended in producing *structure* or *order*. The

total energy content or the specific heat of a material thus provides an indication of the actual state of transformation of the configuration of a group of particles from a state having "structure," that is, a certain degree of order, to a state characteristic of a statistically isotropic (monatomic) gas. The established fact, that the specific heat of molten and of solid metals does not differ appreciably from  $3R$ , whereas the specific heat of metal vapor is about one-half this value, indicates that the amount of

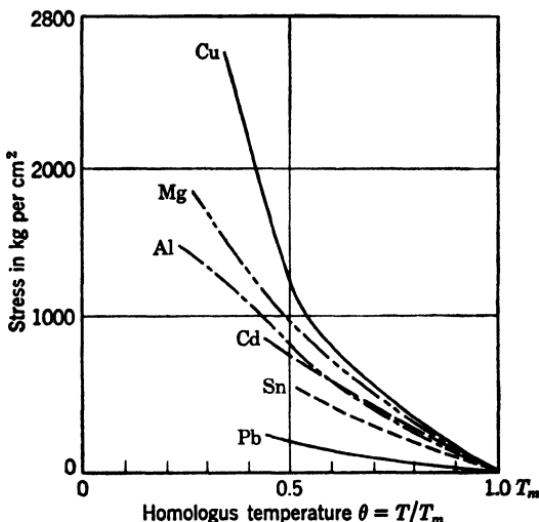


FIG. 10·3 True necking stress  $s = P_{\max}/A$  for various metals at equal homologous temperatures (after Ludwik<sup>10·5</sup>).

"structure" is not considerably changed by the melting and that therefore the internal structure of the melt is very different from statistical isotropy, provided no appreciable density changes have occurred in the transformation from the solid to the liquid state. This is in agreement with the modern theory of liquids according to which a liquid is made up of "flow-units" and "holes."<sup>10·4</sup> Thus the difference between the fluid and the solid state could be visualized as a difference between the mobile, loose packing and the stable dense packing of spherical particles, without appreciable change in the order of magnitude of the forces of interaction.

Because as a result of local thermal oscillations the mean particle energy may be momentarily exceeded, local melting and flow will take place before the temperature of the entire body

has reached the melting point  $T_m$ . A particle will remain stable in its potential trough only as long as the energy impulse of its largest heat oscillation does not attain the amount  $u_0$ . Since by an impulse which exceeds this limiting value, the particle is activated,  $u_0$  is called the *activation, energy*.

The interrelation between activation and melting appears to be the explanation of the so-called *homologous temperatures*, a concept introduced by Ludwik<sup>10-5</sup> in order to compare the mechanical behavior of several metals having different melting points (Fig. 10·3). Ludwik found a fair similarity of mechanical behavior of metals at the same homologous temperatures  $\theta = T/T_m$ . More recent investigations have shown that the concept of homologous temperature is valid only if the compared metals are of similar atomic radius.

## 11. Ordered and Unordered State

According to the previously developed concepts, matter is considered as an aggregation of a large number of discrete particles in space; the behavior of the particles is governed by their content of potential and kinetic energy.

In attempting to classify such assemblies of discrete particles, the principal characteristic is the existence of a certain repetitiveness of pattern, that is, of a periodicity in the geometrical pattern. The state of an assembly of particles in which there is a complete lack of periodicity is usually defined as *unordered* or *disordered*, whereas the existence of a single perfectly repetitive pattern throughout the assembly is the criterion for the state of ideal *order*.

It should be realized, however, that the probability of finding an assembly of interacting particles in a state of complete lack of order is as small as that of finding it in a state of perfect order; perfection, both of order and of disorder, refers to certain arrangements the number of which is very small compared with the almost infinite number of possible arrangements of imperfect order. In the limiting cases atoms are arranged either in a single perfectly repetitive pattern or in a group in which no repetitive pattern, no matter how limited, ever occurs. Both the periodicity and the randomness must be perfect in all three dimensions of space. Particles of real materials exist in neither of the limiting states, but in some intermediate state within the range

of possible combinations of order and disorder (*order-disorder spectrum*). A certain component of order in the arrangement of particles must be expected to exist in any intermediate state between the perfectly ordered ideal crystalline solid and the unordered monatomic gas. It is this component of order that is indicated by the potential energy content of the structure.

There is considerable experimental evidence, particularly from X-ray diffraction patterns of fluids,<sup>11-1</sup> to suggest that particles of amorphous solids and of liquids are not in the same perfectly unordered state that characterizes the monatomic gas, but that only the dimension over which the order extends distinguishes the atomic or molecular arrangement of amorphous materials and of fluids from that of the crystalline solid. Whereas in the crystal the repetitive pattern extends over many thousands of atomic or molecular distances, in the liquid and in the amorphous material (frequently considered to be an *undercooled* liquid since the crystalline material is the only possible type of real solid) there is only local order, which extends over small groups of particles. The local structure in the vicinity of and in relation to an individual particle is very similar to the ordered crystal structure, whereas at a distance from this particle the arrangement with reference to it appears completely random. This arrangement can therefore be considered to exhibit local or *short-range* order but *long-range* disorder.

This can, for instance, be illustrated by the probability distribution curve of the atoms around any considered atom in liquid mercury; this distribution which has been observed by Debye is shown in Fig. 11·1. Mercury does not form molecules, and, hence, the particles in the liquid are atoms. Whereas the abscissas indicate the distance of an atom from the considered atom located at  $\rho = 0$ , the ordinates give the relative probabilities, to a certain scale, of finding an atom at a certain distance  $\rho$ . The distribution function shows that, within small distances from the considered atom, the probable location of neighboring atoms shows a definite periodicity which, however, vanishes for larger

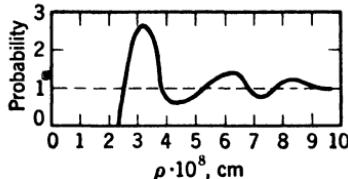


FIG. 11·1 Distribution function of liquid mercury (after Debye).

distances. Thus the probability of finding an atom at a distance 3 is about four times as high as that of finding it at 4, a fact that indicates the existence of ordered patches around each atom. Similar curves have been established for molecular liquids.<sup>11, 12</sup>

The presence of interatomic or intermolecular forces probably causes temporary formation of particles into cohering small ordered groups or *flow units*. These local groups, however, are not correlated with each other and have no sharp boundaries. Therefore, the arrangement in the neighborhood of group boundaries is less regular and also less stable than in the interior of the group, where the forces of interaction between particles are strongest. Consequently, interaction along the group boundaries is broken up more easily than the interaction between the molecules in the interior of the groups. Thus, in moving under the action of an external force, groups sliding over each other will be continuously broken up and re-formed, as the molecules along the circumference of the groups, which are entering and leaving spheres of mutual attraction of different groups, are changing over between various groups. A certain amount of potential energy is thus continuously stored up and dissipated in the course of the formation, motion, and subsequent destruction of the flow units. It is the amount of this energy, expressed by the resistance to the motion, that is, by the *coefficient of viscosity*, that provides an indication of the position of the considered liquid within the order-disorder spectrum.

An ordered or crystalline arrangement of particles in which each particle is surrounded by neighbors according to a definite geometrical pattern repeating itself around every similar particle is called a *space lattice*. In the particular case of all particles being atoms or ions, it is known as a *coordination lattice*, of which thus two types exist; the atomic lattice and the ionic lattice. The number of nearest neighbors in the lattice is the *coordination number*. When the particles in the regular arrangement are molecules, the lattice is a *molecular lattice*.

The grouping of particles according to any possible geometrical pattern that repeats itself periodically in three dimensions results in the creation of differences between certain directions in space. The crystalline substance is therefore intrinsically anisotropic. By rigorous definition only an ideal continuum could be isotropic.

This becomes evident in any attempt to arrange particles according to the definition of isotropy, that is, so that they show identical configurations in all directions. Since this is an impossibility, the isotropy of a discontinuous assembly can only be *statistical*, which means that directional differences in the arrangement around any particular point exist but vary statistically from point to point. Thus, in the classification of the discontinuous arrangement of particles of solid or pseudosolid matter the limiting states are the perfectly ordered anisotropic crystal and the disordered statistically isotropic amorphous substance. In the intermediate states along the order-disorder spectrum the inherent local anisotropy of the discontinuous structure of matter is more important in the solid than in the fluid state; in the solid state the equilibrium position of the particles does not change with time, so that the inherent directional differences in the local arrangement can be smoothed out by "averaging" of behavior over space only, whereas in fluids the "averaging" is done over both space and time.

The arrangement of particles into a perfect three-dimensional space lattice can be considered as homogeneous, if the concept of the equivalence of all points is restricted to the lattice points. Rigorous homogeneity as well as rigorous isotropy would only be possible in the perfect continuum. However, if equivalent coordinate systems can be set up with any of the lattice points as origin, the space lattice may be defined as homogeneous.

Among all possible arrangements of atoms or ions into a coordination lattice, the actually formed lattice will be distinguished by maximum stability, that is, by a minimum content of (negative) potential energy or a maximum value of bond energy, or at least by a tendency to reach such a configuration spontaneously after infinite time.

It has been tacitly assumed, so far, that the atoms in the space lattice are of the same kind. The terms *order* and *disorder* referred to the existence or nonexistence of a periodic arrangement in space of those particles. However, a different definition of order and disorder is used by metallurgists in relation to the ordered arrangement in a space lattice of atoms of different elements such as, for instance, of atoms in metal alloys. In this case either the atoms of the various elements may occupy definite periodic positions within the lattice, forming so-called *super-*

*lattices*, in the lattice points of which one type of atoms is segregated, or the different atoms may be distributed at random over the whole lattice, forming *solid solutions* of one chemical element in the other. The formation of *superlattices* is due to the tendency of the various types of atoms to take up regular positions of minimum energy, thus creating a lattice that consists of several space lattices of various periodicities. The opposing tendency is the result of the thermal oscillations, which tend to produce a random arrangement of particles over the lattice points. The existence of stable homogeneous superlattices, whose potential energies are a minimum, is usually defined as metallurgical *order*; the *solid-solution* random distribution of the various atoms over the lattice as metallurgical *disorder*. Thus, order requires that all particles be in the "right" location, whereas in the condition of disorder or solid solution all particles are located at random. In order to transfer one particle from its "right" to the "wrong" position, energy must be expanded, since the superlattice with all positions "right" represents the state of minimum energy. Methods of statistical mechanics based on the simple consideration of the number of particles in "right" and "wrong" positions may thus be applied to the analysis of the energy changes in the so-called order-disorder transformations in metal alloys.<sup>11-3</sup> These transformations are very important with regard to the mechanical properties of metal alloys produced by various heat treatments. Since the application of elevated temperatures increases the component of disorder as a result of the intensified thermal oscillations, the existence of a superlattice is the less probable, the higher the temperature. On the other hand, the slower the cooling rate after the application of such a temperature, the more probable is the formation of a superlattice. This fact has been confirmed by experiment.<sup>11-4</sup>

## 12. Structural Geometry of the Ordered State

All atoms or ions in coordination lattices are surrounded by neighbors in a regular arrangement, and the arrangement around each atom is identical. Thus, periodic patterns of atomic arrangement are developed in the three dimensions of space.

The ways in which points can be arranged in space in a three-dimensional periodicity, so that each point is surrounded by other points in an identical arrangement, are not unlimited. Studies

of the geometry of such arrangements<sup>12-1</sup> have shown that only 32 different point groups exist forming the 32 *crystal classes*. However, among the 32 classes there are only 14 different translation groups defining *unit cells*, that is, smallest lattice units that fully represent the lattice type, out of which the 14 space lattices can be built. Of these 14 space lattices only 7 have different axes, and these 7 systems of axes are the basis for the classification of crystals into the *triclinic*, *monoclinic*, *rhombic*, *tetragonal*, *rhombohedral*, *hexagonal*, and *cubic* crystal systems. The cubic system has 3 orthogonal axes of equal length, the tetragonal 3 orthogonal axes of which 2 are of equal length, the rhombic 3 orthogonal axes of unequal length, the monoclinic 2 orthogonal and 1 inclined axis, the triclinic system has 3 axes inclined to each other, and the rhombohedral and hexagonal systems have both 3 axes in one plane normal to the fourth axis. Of these 7 systems only the last 3 are particularly important in the study of the behavior of technical metals.

In order to identify a set of parallel planes in a lattice of rectangular axes the so-called *Miller indices* are used.<sup>12-2</sup> They are obtained by finding the intercepts of any of the group of parallel planes with the three orthogonal axes, taking the reciprocals of these values, reducing them to the three smallest integers of the same ratio, and putting them into parentheses (Fig. 12-1). If a plane is parallel to an axis, the corresponding index is zero, since the intercept is infinite. Thus the (111) plane specifies the plane inclined under 45 degrees to all three axes, whereas the index (100) specifies planes perpendicular to the  $x_1$  axis. In order to specify a direction, the coordinates of a point located on a vector in this direction through the origin are specified in the smallest integers of units and put into square brackets. Thus, for instance [100], [010] and [001] define the  $x_1$ ,  $x_2$ , and  $x_3$  axes, respectively, whereas [111] denotes the body diagonal of the unit cell.

The purely geometrical concept of the possible periodic divisions of space leads to the most general classification of crystal-

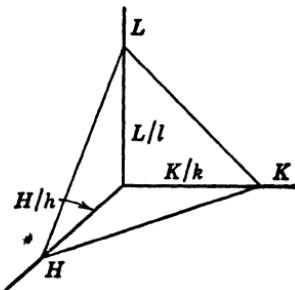


FIG. 12-1 Miller indices  $(h k l)$  of a plane.

line structures. However the discussion of the formation of crystal structures in terms of the packing of the nearly spherical atoms or ions and the possible arrangements of neighboring spheres provides a concept of the crystal structure that is more useful in the interpretation of mechanical behavior of crystalline matter. It also illustrates the effect of differences in atomic radii of the constituent particles on the geometry and the stability of the crystalline structure.

There are a number of possible regular arrangements of "near-neighbor" atoms or ions, that is, of atoms connected by equivalent bonds and pictured as spheres of equal size, around any considered spherical atom or ion:

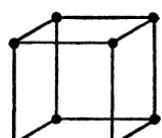
- (a) 2 symmetrical neighbors on a straight line through the atom.
- (b) 3 neighbors in the corners of an equilateral triangle, the center of which is occupied by the atom (Fig. 9-3).
- (c) 4 neighbors in the corners of a tetrahedron, the center of which is occupied by the atom (Fig. 9-4).

The foregoing three arrangements result in the rather loose packing typical of crystal structures formed by covalent bonds. In their formation the so-called  $(8 - N)$  rule,<sup>12-3</sup> which has been derived in Art. 9 from Pauli's exclusion principle, is usually followed. These arrangements are not limited to nonmetals; there are certain metals whose crystal structures are of the 2-, 3-, and 4-neighbor type, respectively, for instance: tellurium with 2 neighbors, antimony with 3 neighbors, and silicon with 4 neighbors (see Art. 9). These metals most closely resemble the

nonmetallic elements which form covalent bonds. The 3-neighbor arrangement is also characteristic for the hexagonal structural formation of a large number of organic compounds (benzene-ring).

FIG. 12-2 Simple cubic lattice. (d) 6 neighbors in the corners of an octahedron, the center of which is occupied by the atom (Fig. 12-2).

This type of packing produces the *simple cubic* lattice. No element is known to crystallize in this form, which is however characteristic for some more complex compounds, for instance the pyrites. It is actually a transition structure between the loosely packed lattices produced by covalent bonds and the closely packed ionic and metal lattices.



(e) 8 neighbors in the corners of a cube, the center of which is occupied by the atom (Fig. 12·3).

This is the *body-centered cubic* structure, in which a number of metals crystallize.

(f) 12 neighbors in the *closest packed* structure of spheres of identical size (Fig. 12·4).

This arrangement leaves the minimum of space between the spheres and may have either cubical symmetry, resulting in the *face-centered cubic* structure, or hexagonal symmetry, resulting in the *hexagonal closest packing* which is slightly less regular than the cubic form of closest packing. The individual atom in the hexagonal closest packing is in contact with a hexagon of spheres arranged in the same plane and with two triangles of spheres above and below.

Closest packing is characteristic of the structure of the majority of metals and a number of ionic crystals. It expresses the tendency of the atoms to form the most stable structure, that is, the structure with the maximum bond energy, positive ions surrounding themselves with the maximum number of negative ions, as in ionic crystals or being crowded together by the attractive forces exerted by the free electrons surrounding them, as in the metal structure.

Actually the 8 neighbor (cubic body-centered) structure is also rather closely packed since, in addition to the 8 nearest neighbors connected by strong bonds, there are 6 neighbors at only slightly larger distances, connected by weaker bonds. For certain metals the total bond energy, and thus the stability of the structure resulting from the 14 bonds of unequal strength, appears to be higher than that produced by the 12 bonds of equal strength of the face-centered group.

Only as long as the structure is made up of spheres of equal size are the cubic and hexagonal closest packings the two alternatives of maximum stability. An infinite variety of arrangements is possible for the closest packing and therefore for maximum stability of spheres of two or a number of different diameters.<sup>8·2</sup> In such structures, mostly metal alloys, intermetallic compounds,

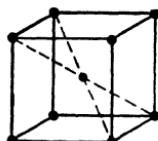


FIG. 12·3  
Body - centered  
cubic lattice.

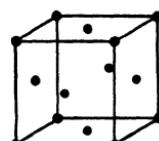


FIG. 12·4  
Face-centered  
cubic lattice.

and ionic crystals, the number of closest neighbors can be increased considerably beyond 12. Several packings are known with as many as 20 neighbors. This refers particularly to compounds of metals with the nonmetallic light atoms of smallest size, such as carbon, nitrogen, and hydrogen. In this case the metal ions are themselves arranged in closest packings and the carbon or nitrogen atoms inserted in the interstitial spaces of the metal lattice (see Art. 8). Such arrangements are called *interstitial solid solutions* (distinguishing them from the *substitutional solid solutions* involving ions of equal or nearly equal size and referred to in Art. 11 in the order-disorder discussion) or interstitial compounds and require usually a slight expansion of the surrounding metal lattice to fit the light atoms into the interstices between the ions. Because of the high density of this arrangement and the large number of bonds formed, these compounds are unusually strong and hard and have extremely high melting points (nitrides, carbides).

In the modern processes of surface hardening of steel parts by *nitriding*, *cyaniding*, and *carburizing*, hard interstitial compounds are formed within a thin surface layer. In addition, since the slight volume expansion associated with the introduction of the carbon or nitrogen atoms into the metal lattice near the surface is prevented by the adjoining metal unaffected by the treatment, relatively high compressive stresses are introduced into the surface layer. Thus, both the high strength and the high endurance of surface-treated steel parts are interpretable in terms of simple changes in the atomic structure conceived as densely packed spheres.

Substances that are built up of identical atoms do not necessarily form one type of structure that is stable under all conditions. On the contrary, many materials, particularly metals, form different structures under different conditions, that is, at different temperatures and under different external pressure, and for different volume concentrations of the various constituent substances in a complex substance. The formation of alternative structural configurations with identical atoms is called *allotropy*.

Allotropic modifications are associated with a different geometrical arrangement of atoms or with a different structure of

molecules, as in the formation of graphite and of diamond out of carbon atoms, of the rhombic and monoclinic modifications of sulphur, or of the cubic (gray) and tetragonal (white) modifications of tin. Either the allotropic modifications are stable at all temperatures or pressures and therefore coexistent, or only one modification is stable within a certain temperature range or at a certain pressure, and transformations from one structure into another occur at definite temperatures or pressures. Such transformations depend very markedly on the rate at which the external conditions are changed.

Allotropic modification of a metal structure can also be produced or affected by deformation producing forced geometrical changes in the existing lattice. Therefore the deformation history of a metal may greatly affect its allotropic transformation under changing temperature. The interrelation between change of temperature and prior deformation in changing the structure and thus the properties of a material are of greatest importance in metallurgical design of alloys, particularly steel.<sup>12-4</sup>

Allotropic transformation is the result of an instability of the existing structure brought about by changing external conditions. The changes extend therefore not only to the form of the lattice, but also to the lattice constants. Thus, in the transformation of iron with increasing temperature from the body-centered cubic  $\alpha$  modification to the face-centered cubic  $\gamma$  modification, the lattice constant, that is, the dimension of the unit cell, increases from 2.9 to 3.6 Å. A unit cell of the face-centered  $\gamma$  iron contains 14 atoms, whereas that of the body-centered  $\alpha$  iron contains only 9 atoms; however the corner atoms are shared by eight cells and the atoms in the faces by two cells. Hence, the atom density per cell is 2 for  $\alpha$  iron and<sup>13</sup> 4 for  $\gamma$  iron. Since the ratio of the volumes of the unit cells is only about 1:1.9, the transformation from  $\gamma$  iron to  $\alpha$  iron is necessarily associated with an abrupt volume expansion accompanied by internal forces or stresses by which the mechanical properties of a polycrystalline aggregate consisting of randomly oriented groups of crystals may be considerably affected.

Only materials that form allotropic modifications under changing conditions, particularly under changing temperature, can be changed by heat treatment. Thus, for instance, pure

copper whose structure is stable at all temperatures, cannot be hardened by a heat treatment, by which the mechanical properties of iron can be very strongly modified.

### 13. Finite Groups. Formation of Real Materials

The energy of a particle located at the surface of a lattice of finite dimensions is higher than the energy of an identical particle located inside the group. This follows from the different equilibrium conditions of particles completely surrounded by interacting particles and of particles in free surfaces, where the one-sided attraction of the particles below the surface must be balanced by tangential forces (surface tension) within the slightly deformed surface and by reduced interatomic distances next and normal to the surface. Hence the total energy of a group of finite number of particles tends to decrease with decreasing relative number of particles at the surface, that is, with increasing size of the group. The most stable group of lowest energy content would therefore be the perfect space lattice of particles extending over the total volume that can be formed.

However, the existence of *structure-sensitive* properties in the so-called "single crystal," which is assumed to be an ideal crystal of finite dimensions, can only be explained by the existence of local inhomogeneities in the space lattice. Although so few in number that they cannot perceptibly influence the *additive* properties (see Art. 3), the effects of the inhomogeneities are sufficient to produce the difference by several orders of magnitude between the observed values of the *constitutive* (structure-sensitive) properties and the values these properties would necessarily attain in the perfect crystal structure. These inhomogeneities in the ordered arrangement could be imperfections within the space lattice, involving isolated particles, such as an occasional missing particle or the occupation of a lattice point by a foreign atom; these chance defects are randomly spaced and they are unavoidable in the building up of any real structure consisting of a large number of elements. Or the inhomogeneities could be concentrated along certain definite planes; this arrangement would produce an approximately periodic subdivision of the primary coordination lattice. Within this subdivision the ordered arrangement of particles would be perfectly homogeneous.

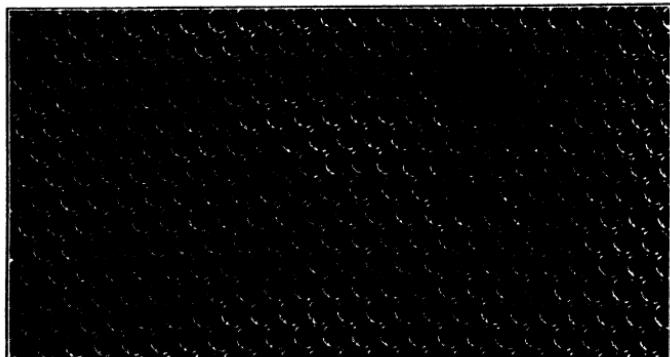
These two assumptions concerning the probable nature of the

imperfections within the ordered structure form the basis of the two large groups of theories concerning the behavior of *real* crystals as distinguished from the perfectly homogeneous *ideal* crystals. The *lattice-defect* and *dislocation* theories are derived from the first assumption; the *mosaic* or *block-structure* theories are based on the second.

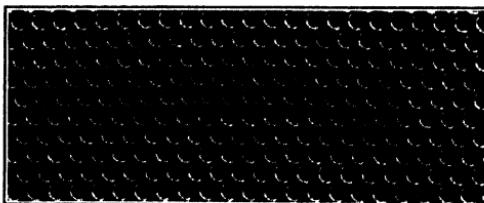
The *lattice-defect* theory has first been developed by Smekal<sup>13-1</sup> who attempted to explain the discrepancy between the atomic bond-strength and the "technical" cohesive strength of solids. According to this theory the existence of an ideal crystal lattice of any size is highly improbable, since there is always a probability, no matter how small, of encountering a defect at any location. The *dislocation* theories amplify the lattice defect theories by introducing a more detailed concept of the "defect," the nature of which in Smekal's theory remains rather vague. The concept of *dislocations* has been introduced by Prandtl<sup>13-2</sup> and developed by Dehlinger<sup>13-3</sup>; it has been applied to the explanation of the discrepancy between the atomic slip resistance and the observed values of the shear resistance in crystals by Taylor,<sup>13-4</sup> Orowan,<sup>13-5</sup> and Polanyi.<sup>13-6</sup> A dislocation is assumed to be a line discontinuity in the atomic lattice such that there is one atom more in the atomic array above this line than below (Fig. 13-1). At the "center of the dislocation" the vertical atomic lines are one-half the atomic distance out of step, whereas they are again in step at considerable distances to the left and right of this center. Thus, the region above the center of the dislocation is under compression; the region below it under tension. Because of the distortion created within the atomic lattice a dislocation contains a certain amount of excess energy so that dislocations constitute points of energy concentration. The assumed density of dislocations varies between  $10^8$  per  $\text{cm}^2$  for annealed and  $10^{12}$  per  $\text{cm}^2$  for heavily cold-worked metal,<sup>13-7</sup> considering planes one atomic distance thick. These figures are equivalent to the volume concentration of defects assumed in the lattice-defect theory of between  $1:10^4$  and  $1:10^8$ , which is also the probability of encountering a defect anywhere in the lattice.

It must be assumed that dislocations are formed both during solidification of the crystal out of the melt and as a result of "unfinished slip," that is, as the result of a slip motion along an

atomic plane which does not pass through the whole crystal but is arrested in its interior by the formation of the dislocation; it is further assumed that the defects associated with the formation of crystals out of the melt are essentially the same type as the dislocations produced by unfinished slip, so that the same mathematical concepts can be applied to the analysis of all dislocations,



*A* Single dislocation.



*B* Row of dislocations.

FIG. 13·1 Soap-bubble models of dislocations in regular array of atoms  
(courtesy Sir Lawrence Bragg<sup>9-8</sup>).

irrespective of their origin. The mathematical theory of dislocations has been developed by Koehler.<sup>13·8</sup>

The *mosaic-structure* theory has first been proposed by Zwicky,<sup>13·9</sup> who attempted to explain the discrepancy between the atomic and the technical cohesive strength of solids by the assumption of "secondary" lattice planes in which the forces of interaction between the atoms are much smaller than in the "primary" lattice planes. Although the theoretical basis of Zwicky's theory was considered untenable, sufficient experimental evidence became available from X-ray investigations of assumedly perfect single crystals to justify the conclusion that

an annealed crystal is composed of *blocks* of perfect atomic arrangement of linear dimension of 1000 to 10,000 Å (one-tenth of a micron to one micron), tilted against each other by about 10 to 15 arc min.<sup>13-10</sup> Because of the perfection of the atomic arrangement within the blocks, their properties could not be structure-sensitive. Contrary to the original mosaic-structure theory, the *block* theory does not postulate the necessity of the existence of blocks or the impossibility of the formation of ideal crystals. It only concludes from the interpretation of the available evidence that practically all assumedly perfect crystals show evidence of a certain degree of disorder, which would suggest the existence of blocks of the afore-mentioned size and relative inclination.

There is, in fact, no real contradiction between the dislocation and the block theory. Both theories can be unified by considering the block boundaries as planes of concentrated dislocations, at least as long as the crystal has not yet undergone any deformation under the action of external loads. Thus, the smallest concentration of dislocations would require that one out of every  $10^4$  atoms be located within a block boundary.

Although the ideal perfectly homogeneous space lattice represents the state of lowest energy, it cannot be assumed that all particles participating in the formation of the lattice will succeed in reaching positions of perfect stability at the moment of solidification, as perfect stability represents a limiting state. The fact that in the case of the existence of *blocks* the material would actually be in a state of higher energy than that associated with perfect stability is irrelevant; it would only indicate that such existence could theoretically not be stable and that the arrangement would therefore tend to change spontaneously, that is, by place changes of individual particles in the course of their thermal oscillations towards the final state of lowest energy; it will however take infinite time to reach this state.

As previously pointed out, there is sufficient experimental evidence that can be interpreted in favor of the existence, in real crystals, of a higher periodicity than that of the ideal homogeneous coordination lattice.<sup>13-10</sup> Such higher periodicity, as expressed by the *block structure*, is due to the fact that group phenomena involving finite groups of particles are operative within the association made up of a practically infinite number

of particles. The size of those groups, which are agglomerations, in ordered condition, of a limited number of particles, is a physical characteristic of the considered material. The experimental evidence tends to show that since groups or blocks are formed during the solidification of the crystalline material out of the melt, solidification is a group phenomenon, that can be described neither in terms of the single particle nor in terms of an infinite number of particles, but only in terms of groups of finite size the formation of which is initiated by *nucleation*.

In the solidification process the formation of groups of finite size apparently precedes the formation of the solid domains; particles arrange themselves into smaller groups according to the configuration of the solid space lattice while still in the liquid state. Conversely, in the melting process such groups persist in formation for several degrees above the melting point. The existence of this *paracrystalline* formation in the melt has been inferred from the persistence, after reheating to several degrees above the melting point, of certain effects of crystal anisotropy observed on previous cooling, if the melt was well protected from outside disturbances.

The particles, which subsequently form the crystal, have considerable freedom to move individually in the melt. As the temperature is lowered, this mobility is gradually restricted, since the tendency to cohere is intensified as a result of the decreasing amplitudes of the thermal oscillations of the particles. Because of the heat fluctuations in the melt solidification does not proceed in accordance with the concept of a crystal built up by the accumulation of particles along the expanding solid boundaries into a three-dimensional continuum. It takes place in a discontinuous manner by the formation, from nuclei, of small solid domains or blocks of particles at locations of the least intense thermal motion. Subsequently, and if given sufficient time, these blocks attempt to form larger units by mutual realignment and by absorption into their group order of individual particles caught between neighboring solidified groups.

If, because of the competing claims as to the order to be followed from adjacent solidified blocks on particles located along the block boundaries, these particles cannot be absorbed into the order of either of the blocks, such particles are kept in positions of excessive potential energy and unstable equilibrium

by a system of forces of interaction connecting these particles simultaneously to the competing ordered block regions. It is doubtful, however, whether such a system has any resemblance to a "dislocation." The position of high energy or *high quantum states* of particles within the block boundaries are characterized by interatomic distances exceeding the stable equilibrium distances in the adjoining block lattices. In the real crystal lattice, consisting of an assembly of small perfectly ordered blocks of imperfect mutual alignment (Fig. 13·2) these positions of high energy are therefore positions of least stability. They are also positions of slight disorder, since they result from the fact that two or a number of competing claims as to the order to be followed have been imposed simultaneously. The character of the substance forming block boundaries is therefore different from its character within the crystallite blocks.

The angular deviation in alignment between adjacent blocks of the same crystal must necessarily be very small, as otherwise the disturbance produced in the whole space lattice would be so considerable as to affect seriously the properties that depend on the lattice as a whole; this, however, does not occur. For the same reason the distortion of the atomic arrangement producing the thickness of the block boundaries cannot extend over more than a few atomic distances. Hence the inhomogeneity in the lattice resulting from the block structure is actually very slight. Its considerable importance in determining the behavior of the lattice under applied forces, however, lies in the fact that these slight irregularities are sufficient to concentrate the initial response to the action of applied forces entirely within the regions where they occur. It is therefore the initial resistance of those boundaries to deformation and separation and not the resistance of the regular lattice that determines the behavior of the real crystal preceding any permanent deformation.

The building up of a perfectly aligned crystal out of small

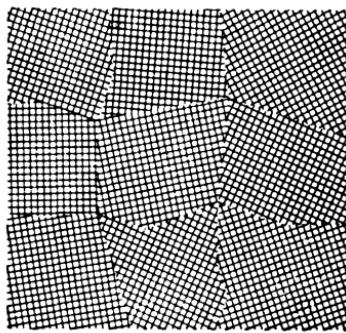


FIG. 13·2 "Block" formation in crystals (mosaic structure).

blocks, theoretically, might be possible if sufficient time were available for such realignment of the blocks to take place before the mutual interference of the blocks formed and the "freezing" of the remaining narrow liquid regions between blocks practically stopped any further free motion. Since the infinitely slow cooling rate necessary for the formation of an ideal crystal out of the blocks cannot practically be imposed, building of the perfectly ordered arrangement of an ideal crystal is impossible. Thus the grade of perfection of the single crystal will depend on the duration of the solidification process.

The formation even of an imperfect single real crystal directly out of the melt can only be achieved under exceptional conditions. Because of the considerable component of "thermal" disorder in the liquid state, the probability of formation of a large number of randomly oriented blocks, which, in the time given, are unable to align themselves into one "single" real crystal, is much more probable. In this case relatively small numbers of blocks of suitable orientation will tend to form "single" crystals of smaller size, the orientation of which will be essentially random. In this formation of statistically isotropic polycrystalline aggregates out of the liquid state, the alignment of blocks forming individual crystals will be the more perfect, and the individual crystals therefore the larger, the slower the cooling process has been; conversely, the crystals will be the smaller, the more rapidly cooling has proceeded. This latter tendency does, however, stop at the limiting order of magnitude of the blocks which constitutes the lower limiting size of the stable associations of particles.

The magnitude of blocks or domains termed *crystallites* by Wood<sup>13-10, 13-11</sup> is of the order of 0.1 cubic micron; it is a characteristic of the material at a certain temperature and depends on its chemical constitution. Thus, a crystallite size of about 0.3 micron for pure iron is reduced to about 0.1 micron by the addition of 0.8 percent carbon.<sup>13-11</sup> Because of the high energy within the interfaces of the blocks, groups of atoms of smaller than crystallite size are thermally unstable and cannot exist. Crystallites therefore cannot be broken up permanently into smaller units since such units are rather rapidly re-formed into the stable crystallite size.

As previously noted, the imperfect alignment of blocks into a

single crystal creates along the block boundaries unstable positions of particles of higher potential energy associated with large interatomic distances. In the same manner the formation of a group of relatively small crystals by solidification causes isolated atoms along the boundary regions between crystals to remain in unstable positions of equilibrium associated with high potential energy of the interatomic forces. These *intercrystalline regions* or *crystal boundaries* differ from the block boundaries by the larger number of particles involved, by the much more pronounced disorder of the atomic structure and by the greater thickness of the disordered regions. Whereas the disorder in the block boundaries is very slight, being due to differences of several minutes in the angular alignment of blocks, the differences in the orientation of the space lattices of adjoining crystals are comparatively large, so that a large number of particles remains within the boundary region, unabsorbed into the order of either neighbor, as long as the number of such particles is not sufficiently large to produce a tendency towards the independent formation of isolated crystallites between the adjoining crystals.

The deformational and strength properties of the polycrystalline aggregate are largely governed by the behavior of the intercrystalline (boundary) regions; those regions are actually responsible for the group behavior of the aggregate.

The initial inhomogeneity of the atomic arrangement within the boundary and block regions due to the varying grain orientation and the imperfect block alignment is further intensified in the course of the solidification process, as a result of differences in the coefficients of thermal expansion in the different directions of the same crystal, as well as for the different constituent elements making up the material. The grain boundaries are thus associated with a high intensity of the interatomic force field, or, as it is usually called, the field of *microstresses* or *textural stresses*. The term *textural stresses*, introduced by Orowan,<sup>13-12</sup> is used in reference to all inhomogeneous fields of interatomic forces arising from the inhomogeneous structure, atomic, molecular, and microscopic, of real materials.

The behavior of the intercrystalline regions is also strongly affected by foreign atoms which, being "strangers" to the groups in formation and thus not easily absorbed, tend to agglomerate along the group boundaries. That this is the case may be

inferred from the observation, that a critical concentration of foreign admixtures (alloys and impurities) exists below which the specific effect of a foreign particle is 10 to 100 times larger than it is above that limit.<sup>13-13</sup> The existence of such a limit is an indication of a sharply differential surface and volume effect of foreign particles. Below the critical limit of concentration the particles agglomerate within the disordered boundary regions; above it they must be absorbed into the volume of the crystals. The critical concentration is very small, but it is exceedingly well defined, particularly at low temperatures; it is reached when the intercrystalline regions or "block-interfaces" are saturated by foreign particles, so that all additional particles are absorbed by the crystals.

The fitting of foreign particles into both the crystal boundaries and the crystal volume is usually associated with the creation of highly localized fields of textural stresses around those particles. These stresses result from the differences between the atomic diameters of the foreign particles and the diameters of the "holes" in the existing atomic arrangement in the grain boundary or within the crystal lattice, into which these particles or groups of particles are being fitted.<sup>13-14</sup>

In the discontinuous assembly of particles the tendency toward an ordered arrangement is essentially produced by electrostatic and valence forces; the tendency toward disorder is due to the thermal motion of the particles. Therefore, the stronger the forces of interaction between the particles, the stronger the tendency to crystal formation, and the less can this tendency be interfered with by thermal motion. A certain balance between the two opposing tendencies is, however, required for the actual formation of crystal structure from the essentially random distribution of particles in the melt, since the thermal agitation of the particles provides their mobility to follow the electrostatic attraction. It is the sharp reduction of this mobility in the cooling process before the electrostatic forces have been able to overcome the disordering tendencies that produces the amorphous structures.

The existence, particularly in metals, of a lower limiting size of ordered crystal arrangement (crystallites), the rapid and discontinuous solidification process leading to block structure, and the paracrystalline formation in the melt are probably interre-

lated phenomena, all of which are due to the combined thermal and time effects governing the transition from the liquid to the solid state. These effects are the result of the momentary relation between the velocities of individual particles in the melt under the action of the ordering (interatomic) and disordering (thermal) influences.

Whereas with decreasing temperature the disordering tendency becomes gradually less violent and the group-forming ordering tendency more pronounced, the reduction of the thermal motion at the same time reduces the mobility of the particles under the influence of the ordering tendency. There can be only a very narrow range of temperatures in which such a relation between the ordering and the disordering tendencies exists that the particles not only have the definite tendency to follow the ordering forces but also are still mobile enough to do so. The maximum distance over which a particle in the melt is free to move within this temperature range probably determines the characteristic size of the crystallite of a particular material. Crystalline materials are thus formed by solidification in a rather discontinuous manner; the melting process is necessarily of similar discontinuous character. The ordering process however, is, not terminated on solidification; it continues for very long periods even in the solid state, as a result of occasional activation by thermal-energy fluctuation of particles in unstable positions. Following the ordering tendency, these particles move or *diffuse* through the lattice to more stable positions of lower energy. In fact, it has been observed that the perfection of the lattice of single crystals increases with time over very long periods.<sup>13-15</sup>

Amorphous materials, on the other hand, are formed by gradual solidification, since the ordering tendency (which is operative before the reduction of temperature prevents further motion of particles) is comparatively weak, being caused by molecular forces. Although a very small-grained polycrystalline aggregate and an amorphous molecular substance show considerable similarities in their mechanical behavior, they still belong to two basically different states. The statistical isotropy of the polycrystalline aggregate and the statistical isotropy of the amorphous substance pertain to different levels of association of particles. Since that part of the behavior determined by the character of

the element on each level is invariable, deformational behavior of a crystalline aggregate, however small-grained, will retain its characteristics to such an extent as to be distinctly different from that of an intrinsically amorphous substance, the statistical isotropy of which is created at a level of association at which the element is the atom or the small molecule.

It has been mentioned previously that transformation from the liquid into the solid state can be achieved not only by cooling but also by polymerization. The solid chain molecular structure formed by polymerization either may be so thoroughly inter-linked that its rigidity is high and, within a certain range, practically unaffected by temperature, or it may consist of linear chains with a comparatively small number of weak cross links and therefore manifest high deformability and temperature sensitivity. In the formation of high polymers, it is the size and shape of the reacting molecules that determines the structure, and thus the mechanical properties.

A few materials can exist in either the ordered or the unordered state, quartz being an example. In this case the amorphous state is not entirely stable; it is generally possible to transform the amorphous into the crystalline state by long periods of moderate heating, during which the mobility of the particles is so intensified that, at least over long periods of time, they tend to follow the ordering forces. Such transformations can be speeded up by the introduction of special ordering effects, such as strains, by which the ordering forces are intensified. In those materials a very slow cooling process increases the probability of achieving ordered formation, whereas rapid cooling increases the probability that the state of disorder existing in the melt will be preserved in the solid. This preservation of the disorder of the melt which results in the *undercooled liquid* formation of the glasses depends upon the existence in the melt of groups of polymerized cations. These groups, the elements of which form in the solid state ionic bonds with the anions present, must be so large or so irregular that they cannot be simply added to a growing crystal lattice. Thus the glass-forming tendency is intensified by increasing dimensions or irregularity of the polymerized groups and vanishes when such groups do not exist.

Simultaneous existence in either of the two states is a phenomenon occurring also in the formation of polycrystalline aggre-

gates out of the melt; however, in this process the unordered state, which is developed in the interstitial regions between the crystals, can survive only as a result of the conflicting ordering claims concerning the position of individual particles in the liquid held between the solidified crystals and the extent of disorder is restricted to layers of thickness of a few atomic distances.

#### 14. Microscopic and Macroscopic Structure

On the microscopic scale of observation, the structure of engineering materials is either truly amorphous or "technically amorphous" or polyerystalline.

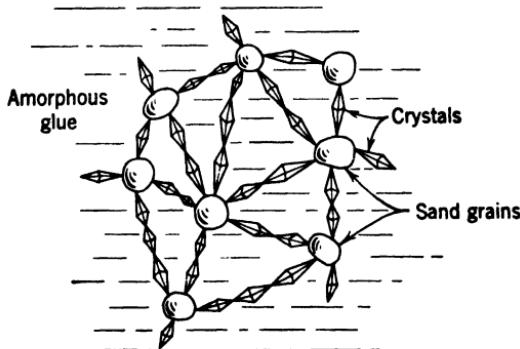


FIG. 14·1 Structure of cement-sand mortar.

Truly amorphous structures are those of materials of the *undercooled liquid* type such as glass, some asphalts, and some nonhardening resins. Their mechanical behavior is determined by the interaction of molecules.

All "technically amorphous" materials are two- or multiple-phase systems, consisting of a continuous spatial network of a certain material, the spaces of which are filled by other materials in molecular or microscopic distribution. Either the network is a statistically isotropic rigid skeleton sometimes crystallized, held together by primary bonds, such as the silicate skeleton of the ceramic materials, the calcium-aluminum skeleton of Portland cement or materials containing additional microscopic particles, such as the cement-sand mixtures (Fig. 14·1); or it may be a flexible network of very large molecules, which show preferred orientation leading to a definitely anisotropic inhomogeneous structure, as in textile fibers.

Substances filling the spaces within the network range from molecularly distributed water in clay, ceramic materials, and gelatin, to oily or very viscous gluelike liquid phases such as are present in asphalts, synthetic resins, and hardened cement. In the analysis of the mechanical behavior of the "technically amorphous" materials, both the interactions on the molecular scale and the behavior resulting from the presence of two or more constituent phases of different molecular structure have to be considered. The order of magnitude of the structural assembly is intermediate between the molecular and the microscopic; it depends on the actual dimensions of the constituent particles.

Statistically isotropic polycrystalline materials such as minerals and metals consist of an assembly of crystal grains of an order of magnitude ranging between several microns and several millimeters. The initial sizes of the crystal grains within the aggregate depend primarily on the duration of the solidification process; they may, however, be subsequently modified by temperature and deformation. Since the stable size of a crystal grain is a function of temperature, there probably exists an *equilibrium state* of distribution of crystal sizes within the aggregate, toward which the actual distribution tends with time.

The mechanical behavior of polycrystalline aggregates is built up from that of the single grains under the mutual restrictions imposed on their deformation by the existence of adjacent grains and of grain boundaries. The structure of the polycrystal can be considered as a two-phase system consisting of randomly oriented crystalline regions forming a continuous spatial aggregate held together along the boundaries by a very thin layer of unordered material.

Some technically important materials cannot be considered homogeneous, even on a microscopic scale, but are built up of two or more different materials, each one providing certain properties. Such materials are either statistically isotropic or definitely anisotropic.

The majority of materials of this type, which may be termed *macroscopically heterogeneous*, are two-phase systems, containing one solid phase and one fluid (viscous) phase. Frequently this viscous phase is itself a two- or multiple-phase substance, such as cement, certain types of asphalts, or glues and high-polymer bonding materials.

The most important examples of macroscopically heterogeneous materials are concrete (consisting of a rigid skeleton formed by the aggregate, filled by the highly viscous cement paste), technical asphalt mixtures (where the viscous phase is formed by the bituminous material, the solid phase by the stone filler), and the laminated plastics (built up of wood, glass, or high-polymer layers bonded by adhesives). The mechanical behavior of heterogeneous materials is determined by the proportion of the different materials present. In statistically isotropic materials, such as concrete, the characteristic deformational behavior will be that of the phase that is continuously distributed: when the solid skeleton is continuous, the mechanical behavior is that of a solid, as in a rather lean concrete; when the viscous phase is continuous, the behavior is essentially that of a fluid, as in a rich asphalt mixture or a very rich concrete. By the re-establishment of the continuity of the solid phase in the course of the deformation (interlocking of stones in concrete), an initially viscous behavior may be blocked and transformed into that of a solid.

Evidently the forces of interaction between the different phases are as important a factor in determining the mechanical properties of the material as are the forces of interaction within each phase. Forces between phases are the surface forces building up the adhesion between the solid particles and the viscous phase, or between the bonded material and the adhesive. The relative influence of the two phases is therefore determined not only by their proportion in volume, but also by the relation between the internal forces within each phase and the forces of interaction between the phases. Thus the performance of laminates is essentially governed by the relative resistance to deformation and fracture of the bonded sheets and of the adhesive, and not by the strength of the bonded material alone.

### References

- 9.1 F. O. RICE and E. TELLER, *The Structure of Matter*, John Wiley & Sons, New York (1949) 112.
- 9.2 W. SHOCKLEY, *J. Applied Phys.* **10** (1939) 553.
- 9.3 L. BRAGG, *North East Coast Inst. Engrs. & Shipbuilders* **60** (1943) 299.
- 9.4 L. BRAGG and J. F. NYE, *Proc. Roy. Soc. A* **190** (1947) 474.
- 9.4 H. GRIMM and H. WOLFF, *Handbuch d. Physik* (Geiger-Scheel) **24/2** (1933) 958.
- 9.5 R. HOUWINK, *Elasticity, Plasticity, and Structure of Matter*, Cambridge Univ. Press (1937) 23.

9·6 F. SEITZ, *The Modern Theory of Solids*, McGraw-Hill Book Co., New York (1940) 380.

10·1 C. ZWIKKER, *Technische Physik d. Werkstoffe*, J. Springer, Berlin (1942) 85.

10·2 C. SCHAEFER and F. MATOSSI, *Das ultrarote Spektrum*, J. Springer, Berlin (1930).

10·3 R. C. TOLMAN, *Principles of Statistical Mechanics*, Oxford Univ. Press (1938) 93.

10·4 S. GLASSTONE, K. J. LAIDLER, and H. EYRING, *The Theory of Rate Processes*, McGraw-Hill Book Co., New York (1941).

10·5 P. LUDWIK, *Z. Ver. deut. Ing.* **59** (1915) 657; *Stahl u. Eisen* **35** (1915) 1188.

11·1 J. G. KIRKWOOD, "The Structure of Liquids," in *Science in Progress*, Yale Univ. Press, New Haven (1942) 208.

11·2 T. ALFREY, *Mechanical Behavior of High Polymers*, Interscience Pub., New York (1948) 69.

11·3 F. C. NIX and W. SHOCKLEY, *Rev. Modern Phys.* **10** (1938) 1.  
W. BOAS, *Physics of Metals and Alloys*, John Wiley & Sons, New York (1947) 145.

11·4 A. GUINIER, *Radiocrystallographie*, Dunod, Paris (1945) 190.

12·1 W. H. ZACHARIASEN, *Theory of X-Ray Diffraction in Crystals*, John Wiley & Sons, New York (1945) 24-81.

12·2 W. BOAS, *op. cit.*, 5.

12·3 *Ibid.*, 46.

12·4 J. H. HOLLOMON and L. D. JAFFE, *Ferrous Metallurgical Design*, John Wiley & Sons, New York (1947) 46.

13·1 A. SMEKAL, *Physik Z.* **34** (1933) 633.

13·2 L. PRANDTL, *Z. angew. Math. & Mech.* **8** (1928) 85.

13·3 U. DEHLINGER, *Ann. Physik* **2** (1929) 749.

13·4 G. I. TAYLOR, *Proc. Roy. Soc. A* **145** (1934) 362.

13·5 E. OROWAN, *Z. Physik* **89** (1934) 634.

13·6 M. POLANYI, *Z. Physik* **89** (1934) 660.

13·7 F. SEITZ and T. A. READ, *J. Applied Phys.* **12** (1941) 178.

13·8 J. S. KOEHLER, *Phys. Rev.* **60** (1941) 397.

13·9 F. ZWICKY, *Helv. Phys. Acta* **3** (1930) 269, 460.

13·10 A. GOETZ, *Proc. Intern. Conf. Phys.* **2** London (1934) 62.  
N. P. GOSS, *Trans. Am. Soc. Metals* **24** (1936) 967.  
P. LACOMBE, *Conf. Strength of Solids*, Bristol 1947 *Phys. Soc. London* (1948) 91.

13·11 W. A. WOOD and W. A. RACHINGER, *J. Inst. Metals* **75** (1949) 571.

13·12 E. OROWAN, *Symposium on Internal Stresses*, *Inst. Metals London* (1948) 47.

13·13 G. MASING, *Handbuch der Metallphysik* **4(2)** Berlin (1941) 146.

13·14 F. R. N. NABARRO, *Symposium on Internal Stresses*, *op. cit.*, 237.

13·15 A. GUINIER, *op. cit.* 208.

## STRUCTURAL THEORIES OF DEFORMATION

**15. Statistical Aspect of the Behavior of an Assembly of Particles**

The change of mechanical *state* of a material is the expression of changes within its structure or in the group pattern formed by the structural elements at the considered level of aggregation. Changes of state which are reflected in the shape of the stress-strain or of the strain-time diagrams observed in engineering tests are either the result of the momentary action of external forces or independent of this action, being the result of the "spontaneous," thermal motion of the atoms or molecules. Thermal motion, however, can lead to "spontaneous" changes in the structural pattern only when the existing pattern is not perfectly stable; such changes would therefore not be possible either within the ideal crystal lattice or within materials of perfectly unordered, that is, statistically isotropic arrangement of particles, since both types of grouping of particles represent, at different levels of temperature, limiting conditions of perfect stability. The possibility of a "spontaneous" change of state within a material would therefore depend on the existence of a certain amount of disorder within the essentially ordered structure, or of a certain amount of order within the essentially unordered structure; since the energy content of the existing structural pattern would thus not be a minimum, a tendency of "spontaneous" change of the grouping of particles toward a more stable arrangement of lower energy would establish itself. The stability of the structural pattern, however, may be considerably changed in the course of changes produced by external forces.

alone; a certain interrelation must therefore always be expected to exist between the force-dependent and the "spontaneous" changes of state in that the momentary intensity or rate of the "spontaneous" change of state will be a function of the preceding force-dependent change if such change has been associated with identifiable changes in the structure of the material and thus in its thermal stability.

In any case of change of the atomic or molecular pattern, particles have to be loosened from their equilibrium position in the structure, in which they are held by the bond forces, and transferred into a new position. The energy required for this local loosening of the structure by thermal *activation* of the particles must be at least equal to the *activation energy*. This energy  $Q = qN$  cal per mole, where  $u_0 = q$  is the activation energy per particle and  $N$  the Avogadro number.

The sources of supply for the energy necessary to overcome the potential barrier of the interacting forces, which prevents the escape of a particle from its equilibrium position, are the external forces and the thermal oscillations of the particles. Spontaneous change of energy and position of a single particle within a group thus depends on the local accumulation, during the time necessary for the "escape" or *activation* of the particle, of sufficient thermal energy to overcome the potential barrier represented by the activation energy. The rearrangement, by such changes, of the position of particles in a group thus depends on the distribution, at any moment, of the heat energy within the group.

The problem of evaluating the frequency (or the probability) of any particle in the group attaining a state of energy in which it is free to migrate from its equilibrium position is a problem of classical statistical mechanics. It is the problem of finding, among all possible configurations or "patterns" of an assembly of particles, possessing a finite total amount of energy and enclosed within an isolated part of space, that pattern which has the highest probability. This pattern, therefore, will not only occur most frequently but, if disturbed, will also spontaneously re-establish itself because, being more probable than any other configuration, it is also more probable than the disturbed configuration. It represents therefore the configuration with the highest stability.

In order to apply probability considerations to the distribution of particles in space and to the energy distribution over the particles, it must be assumed that there is no coupling of the states, that is, of the distributions at different times and that each state represents a mean value over a considerable period of time, which is independent of the selected period and corresponds to the mean value deducted from the most probable state by considerations of probability. This so-called *quasiergodic* hypothesis asserts, in short, that, whatever the initial state, a stationary state which is the state of greatest probability is attained in time.

By computing by statistical methods the most probable geometrical distribution of particles within a finite volume, divided into cells of different size, it can be easily shown that the number of particles in the individual cell is proportional to the size of the cell; the particles in the stable state are therefore uniformly distributed. It is towards this distribution that any random distribution will tend in time. Similar considerations applied to the distribution of a finite amount of energy  $E$  over the particles lead to a different condition of stability because of the subsidiary condition that the total amount of the energy of the particles is finite and that therefore  $E$  must be equal to the sum of all component amounts of energy  $\epsilon_i$  associated with the individual particles.<sup>15·1</sup> If  $N_i$  denotes the number of particles in the energy state  $\epsilon_i$ , the total energy,

$$E = \sum^i N_i \epsilon_i \quad (15 \cdot 1)$$

where a definite distribution of energy states  $\epsilon_i$  is described by the number  $N_1, N_2 \dots N_i \dots$  etc., of particles in each state. Evidently

$$\Sigma N_i = N \quad (15 \cdot 2)$$

where  $N$  denotes the total number of particles. The fraction of particles occurring in a state of energy characterized by the subscript  $i$  defines the probability  $p_i$  of finding an individual particle in this state; hence,

$$p_i = \frac{N_i}{N} \quad \text{and, because of eq. } 15 \cdot 2, \quad \sum p_i = 1 \quad (15 \cdot 3)$$

If we introduce, instead of the number  $N_i$ , the fraction  $dN$  of particles falling within an infinitesimal range of energy  $\epsilon$ , the energy increase per particle, eq. 15·3 becomes

$$p = \frac{dN}{N} \quad (15\cdot 4)$$

The energy distribution of highest stability and thus of maximum probability under the auxiliary conditions 15·1 and 15·2 can be derived by applying the calculus of variations to the probabilities of the possible energy distributions. The problem has been dealt with by Boltzmann and by Gibbs, who has called the resulting distribution a "canonical distribution," in which the number of particles in various states of energy  $i$  is given by the equation:

$$N_i = Ne^{\frac{\phi - \epsilon_i}{\theta}} \quad (15\cdot 5)$$

The values  $\phi$  and  $\theta$  in this expression are determined in accordance with the subordinate conditions 15·1 and 15·2. It can be shown that the *modulus of the energy distribution*  $\theta$  is the mean energy  $\epsilon_0$  of the oscillating particles for each degree of freedom, which is  $\epsilon_0 = kT$ , whereas  $\phi$  is a constant, characteristic of the potential energy of the distribution of particles.

The probability  $p_i$  of finding an individual particle in the state of energy  $\epsilon_i$  can thus be expressed in the form,

$$p_i = e^{\frac{\phi - \epsilon_i}{\theta}} = e^{\phi/kT} \cdot e^{-\epsilon_i/kT} = Ce^{-\epsilon_i/kT} \quad (15\cdot 6)$$

or, according to eq. 15·4,

$$p = \frac{dN}{N} = Ce^{-\epsilon_i/kT} \quad (15\cdot 7)$$

Equation 15·6 is the basic equation governing the energy distribution within stable systems of discrete particles and is usually referred to as the Maxwell-Boltzmann<sup>15·1</sup> distribution. It is widely used in the analysis of all processes for the initiation or the progress of which the necessary energy has to be supplied by fluctuations of thermal energy, since it relates the thermal energy  $\epsilon$ ; or, if written in the form of eq. 15·7, the energy increase  $\epsilon$  per particle to the expected probability of attaining it. This probability in turn defines the expected number of times per

second that any particle will experience an energy fluctuation  $\epsilon$ , if the frequency  $\nu$  of the energy oscillations is known. With  $\epsilon = u_0 = q$  for the atom or with  $\epsilon = Nq = Q$  for the mole, eq. 15·7 expresses the probability of *activation*. If  $\epsilon$  or  $Q$  is expressed in thermal units, Boltzmann's constant  $k$  is replaced by the gas constant  $R$ .

A formula of similar type has first been proposed by the chemist Arrhenius to describe the rate at which processes of chemical reaction take place at various temperatures; in this form it is being widely used in the theory of the so-called "rate processes."<sup>15·2</sup>

The expression for the mean energy  $\epsilon_0$  of an oscillating particle as obtained from eq. 15·6,

$$\epsilon_0 = \frac{E}{N} = \sum^i \epsilon_i p_i = C \sum^i \epsilon_i e^{-\epsilon_i/kT} = C \int_0^\infty \epsilon e^{-\epsilon/kT} \cdot d\epsilon = kT \quad (15 \cdot 8)$$

confirms the assumption introduced previously.

Equation 15·7 has been derived on the basis of classical (non-quantum) statistical mechanics. However, the processes both of energy absorption by which the particles reach the activated state and of energy emission associated with the subsequent adjustment in finding a new position of higher stability are "quantized" and can take place only if the absorbed and emitted energies are multiple integers of the "quantum" of the particle; they can be described by classical mechanics only if the amounts of energy involved are so large that the increase in the successive energy levels by quantum steps can be regarded as practically continuous. Equations 15·6 and 15·7 may therefore be assumed to describe adequately the energy distribution over the particles occurring at temperatures at which the mean energy of the individual particle is relatively high. To describe changes at temperatures, however, where the quantum effect is pronounced, the Maxwell-Boltzmann energy distribution must be modified to give a distribution law consistent with the possible quantum states. According to this modification<sup>15·3</sup> the probability that a particle will find itself in the quantum state  $n$  is

$$p_n = \frac{e^{-\epsilon_n/kT}}{\sum_n e^{-\epsilon_n/kT}} \quad (15 \cdot 9)$$

where, according to Planck's relation  $\epsilon_n = nh\nu$ . According to the modified eq. 15·8,

$$\epsilon_0 = \frac{h\nu(e^{-h\nu/kT} + 2e^{-2h\nu/kT} + \dots)}{(1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \dots)} = kT \cdot \frac{\frac{h\nu}{kT}}{e^{h\nu/kT} - 1}$$

$$\therefore = kTP(x) \quad (15 \cdot 10)$$

Equation 15·10 is Planck's expression for the average energy content of a quantized oscillator. That the difference between the energy content of a Planck oscillator and that of a classical (Newtonian) oscillator  $\epsilon_0 = kT$  is considerable can be shown by plotting the *Planck function*  $P(x) = x/(e^x - 1)$  where  $x = h\nu/kT$  (Fig. 15·1). This diagram shows that the mean thermal energy of Planck oscillators is considerably lower for small values of  $T$  (that is, small values of  $1/x$ ) than the energy of a classical

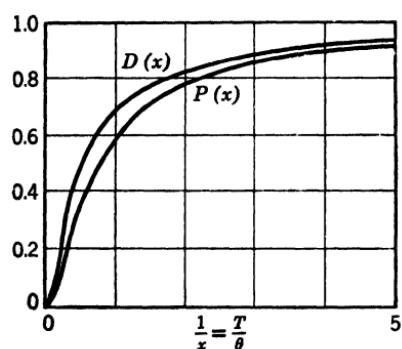


FIG. 15·1 Planck's function  $P(x)$  and Debye's function  $D(x)$ .

oscillator  $kT$ . For large values of  $T$  however the mean energy of the Planck oscillator approaches the classical value  $\epsilon_0 = kT$ , as  $P(x)$  approaches unity.

The difference between the mean energy of the classical and of the Planck oscillator is important with regard to the mechanical behavior of materials at low temperatures. Thus, for instance, the separation of a cohesive bond which initiates fracture within the atomic structure is a "stabilizing" process, that is, a process associated with a decrease of potential energy (since otherwise it would not take place); it occurs only if the instability of the momentary position produced by the action of forces cannot be resolved by an alternative stabilizing process, such as the "spontaneous" migration of the unstable particles to a more stable position. Because of the considerably lower mean energy and because of the quantization of both the activation and the subsequent stabilization process, the possibilities of place change

of particles are less numerous, and migration by activation therefore less probable than they would be according to the classical relation. With decreasing temperature, the alternative stabilizing mechanism of bond disruption becomes, therefore, relatively more probable and thus more frequent.

So far the energy considerations apply to a one-dimensional oscillator. In a solid every particle can be regarded as a three-dimensional oscillator. Hence, by the rules of classical statistics the mean energy of the particle is  $3kT$  or, for one mole of the substance,  $E_0 = 3NkT = 3RT$ . Since the gas constant  $R$  is approximately 2 cal per degree, the specific heat  $c = \partial E_0 / \partial T = 3R$  is about 6 cal per degree (see Art. 10). Experiments, however, show considerable deviation from this figure, which are the greater, the more firmly the particles are held in their equilibrium positions, that is, the stronger the interatomic forces and the lower the temperature. It has been found, for instance, that the specific heat per mole for diamond at room temperature is only about 1 cal per degree. The deviations from the classical figure are due to the quantization of the vibrational energy. Hence, the mean thermal energy of the oscillators per mole of the substance, and thus the true specific heat, can be obtained by integrating eq. 15·10 over the  $3N$  degrees of freedom of the  $N$  oscillators, considering that the  $N$  oscillators are coupled and cannot all perform oscillations of the same frequency. Therefore,<sup>15·4</sup>

$$E_0 = kT \int_0^{3N} P(x) ds = 3RT D(x) \quad (15\cdot11)$$

where  $ds$  denotes the number of oscillations per unit volume between the frequency  $\nu$  and  $(\nu + d\nu)$  and  $D(x)$  represents the so-called *Debye function* which, like the Planck function, approaches unity for small value of  $x$  (Fig. 15·1). The absolute temperature  $T = \Theta$  associated with  $x = 1$  is known as the *Debye* or *characteristic* temperature. It indicates the temperature below which the material no longer behaves even nearly in accordance with classical concepts. If this temperature appreciably exceeds room temperature, quantum considerations may influence the mechanical behavior of the material even at room temperature. Table 15·1 shows approximate values of the Debye temperatures of several elements; this temperature can be computed by differ-

ent methods which lead to slightly different results; the given values represent averages.

TABLE 15·1

	$\Theta$ in °K		$\Theta$ in °K
Pb	90	Cu	315
Na	150	Al	395
Zn	210	Fe	450
Ag	215	Be	1000
Mg	290	C (diamond)	1860

According to Table 15·1, it appears that quantum effects with regard to large-scale (engineering) behavior are negligible for lead, zinc, and silver only; for other metals these effects may become important in the atomic interpretation of results of mechanical tests performed at low temperatures.

The maximum mean frequency of thermal oscillations is related to the Debye temperature by the relation,

$$\nu_{0\max} = \frac{k\Theta}{h} \quad (15\cdot12)$$

Below this temperature the frequency  $\nu_0$  may be assumed to vary roughly with  $kT/h$ .

## 16. "Spontaneous" Changes of State.

All changes of the geometrical pattern of the internal structure of a material resulting from changes of the energy pattern that are "spontaneous", that is, independent of the momentary value of the applied external forces, are produced by the tendency of the structural elements to assume finally a position or a shape of maximum stability. This tendency can assert itself in spite of the solidification of the material, because the incessant thermal activation of the atoms and molecules oscillating about their positions of equilibrium produces a finite probability of occurrence of a thermal-energy fluctuation sufficient to loosen individual particles from their momentary equilibrium position,

enabling them to move into more stable locations associated with lower potential energy. The intensity of the resulting migration of particles within the solid structure determines the characteristic *diffusion rates* of particles of various elements. The diffusion process, the progress of which is essentially not different from a true chemical reaction, is thus necessarily governed by the probability of activation of individual particles, and the momentary diffusion rate  $D$  can therefore be expressed by a function of the form of eq. 15·7,

$$D = A e^{-Q/RT} \quad (16\cdot 1)$$

where  $A$  denotes a structure-dependent constant and  $Q$  the activation energy or the *loosening* energy of the structure. It is usually assumed that both constants depend only on the existing structure and are independent of temperature, the temperature dependence being expressed entirely by the exponential function. This assumption is only approximately true since the constant  $A$  is a linear function of the frequency of the thermal oscillations of the particles which is temperature-dependent, and the constant  $Q$  is necessarily a function of the volume expansion of the structure produced by temperature: this latter temperature effect is, however, not very significant in the solid state.

Equation 15·7 only expresses the *probability* of occurrence of an energy fluctuation sufficient to enable a particle to migrate from its present position. The *rate* of migration is defined by the *expected number of particles* migrating per second into alternative positions and is therefore given as the number  $\nu$  of successful trials in the number of total attempts to reach the activation level which is represented by the mean frequency  $\nu_0$  of the thermal oscillations of the particles. Since the frequency of thermal oscillations within the range of "nonclassical" behavior ( $T < \Theta$ ) of the material can be expressed by (see Art. 15)

$$\nu_0 = \text{const} \frac{kT}{h} \quad (16\cdot 2)$$

eq. 16·1 can be written in the form,

$$D = A_1 \frac{kT}{h} e^{-Q/RT} \quad (16\cdot 3)$$

which is the form known in rate processes. The new constant

$A_1$  is now dependent on structure only. For  $T > \Theta$  the factor  $kT/h$  in this formula is replaced by the constant  $v_{0\max}$ , according to eq. 15.12.

Both constants  $A_1$  and  $Q$  will necessarily change with changing structure, no matter whether the structural changes are caused by external forces or by the "spontaneous" temperature effects themselves. The effect of such changes is different, however, on either of the constants: it has been found<sup>16.1</sup> that  $Q$  is apparently more affected by changes in the atomic structure and its inhomogeneities whereas  $A_1$  is more sensitive to changes in the microscopic structure. Thus, changes in the size of crystals in polycrystalline aggregates must be expected probably to influence  $A_1$  considerably more than  $Q$ .

Migration of particles within the solid structure to more stable positions as a consequence of their momentary unstable position and of their activation by thermal-energy fluctuations is usually referred to as *diffusion* only as long as the number of particles involved in the process is relatively small. If the particles migrating are of the same nature as the particles of the solid structure, the process is defined as *self-diffusion*. As the number of particles migrating increases, the self-diffusion process is gradually transformed into a process of large-scale breakup and re-formation of the existing structure as, for instance, *recrystallization* in polycrystalline metals. If the intensified diffusion process involves particles of various elements, the resulting structural changes are considered as *phase transformations*. If, after diffusion, the chemically different particles aggregate within preferred locations, usually block and grain boundaries, the process is referred to as *precipitation*. In both the phase transformation and the precipitation processes true chemical reactions may occur by the formation of new molecules. Hence, the transformation from either process to a genuine chemical reaction is gradual, and the classification of a process as precipitation, phase transformation, or chemical reaction may, to a certain extent, become arbitrary.

Diffusion of particles within a crystal lattice is associated with the coexistence of particles of widely different size, with the existence of imperfections in the lattice, or with such distribution of atoms of different elements that the energy of the real lattice is raised above the minimum associated with the distribution of

atoms in the lattice of an ideal crystal. There are in a crystal lattice three possible types of migration or of place change:

1. Mutual place change of two particles of different chemical character without structural change, for instance the place change of particles of two metals in solid solution in the process of establishing a superlattice. This process involves the simultaneous activation of two particles; it is usually called *substitutional diffusion*.

2. Diffusion of small particles through the interstices between the large particles forming the lattice. This process, called *interstitial diffusion*, mostly involves atoms of C, N or O which, because of their small atomic radii are the most mobile particles within the lattice (see Art. 8). Since interstitial diffusion depends on the activation of individual particles, it is more probable and therefore more rapid than substitutional diffusion.

3. Diffusion into vacant lattice points which usually takes place along *internal surfaces*, that is, along planes of distorted lattice arrangement in which the interference with the movement by the other particles is relatively small, such as block boundaries. Because of the reduced interference this is the most rapid type of crystalline diffusion.

As a result of the disturbed equilibrium of particles within the surface of a crystal lattice, the stability of their position is lower than the stability of particles in the interior. Hence surfaces are thermally less stable, and surface particles more easily activated and more freely mobile than particles in the interior of the lattice. Therefore diffusion processes involving surface particles are more rapid than diffusion of particles from the interior. The total rate of diffusion in a crystal lattice or a polycrystal must thus be expected to increase with increasing surface or interface areas, that is, with decreasing size of individual crystals. This fact explains the observed inverse relation between the intensity of various diffusion processes and crystal size in polycrystalline metals.

Because of the fluctuation of the thermal energy of the oscillating particles the internal structure of every material must be considered as being continually involved in a process of spontaneous change which, as the structural pattern approaches conditions of maximum stability, proceeds at a decreasing rate, becoming finally so slow as to be no longer perceptible. A

multitude of technologically most important phenomena such as polymorphous transformations, aging, precipitation, oxidation, metallurgical order-disorder transformations, recovery, and recrystallization are results of diffusion processes. Although the activation energies and the constants of proportionality differ widely in the different processes, changing even in the course of the same process, the temperature dependence of the rate at which these processes take place is invariably governed by eq. 16·1 and, for low temperatures, eq. 16·3.

The changes produced in the atomic or molecular structure of the material by diffusion processes may considerably affect the subsequent behavior of the material under the action of external forces, even if the diffusion process has not resulted in a phase transformation, or in a definite change of the size or shape of the larger structural elements, as in recrystallization or polymerization processes. By diffusion of particles *into* certain positions within the structure of crystals or polycrystalline aggregates or *from* such positions, elastic energy is either stored up within the matrix surrounding those particles or released. In either case the response to a subsequently applied force is necessarily changed. The stored-up elastic energy can be considered as the strain energy created in the lattice by expansion or contraction due to the migrating atoms attempting to fit into the positions between the stable atoms. If the considered body is free of stresses produced by external loads, this energy is due entirely to the elastic resistance of the matrix surrounding the migrating atoms. If, however, the body is subjected to external forces, the atoms in expanding the lattice do work against the forces. The stored-up elastic energy will therefore vary with the position of the atom in a nonhomogeneous stress field, and unstable atoms will migrate in such a way that a maximum amount of elastic energy is dissipated. Thus the diffusion processes going on in the material do not only influence its subsequent response to external forces but are themselves affected by the simultaneously acting stress field and, still more, by irrecoverable changes in the structure produced by such stresses. Thus, the interrelation between changes of state due to primarily force-independent changes of the structure and changes of state produced by the action of external forces is necessarily rather complex, and it must be expected that the velocity of the forced

change of state will not be independent of the rate of diffusion of the particles whose migration to new positions produces or releases stored-up elastic energy within the material. This conclusion is borne out by a number of investigations<sup>16-2</sup> in which definite relations could be established, for instance, between the velocity of the forced deformation of steel and the diffusion rate of particles of certain elements, such as carbon and nitrogen within the iron lattice.

The significance of the diffusion rate of such foreign particles may become especially pronounced if deformations proceed under sustained load, as in creep tests, or under very small stress rates. In such cases a certain load level will produce an immediate elastic response only but no irrecoverable deformation if applied rapidly enough. If the load is sustained, however, a rather sudden irrecoverable deformation may start after a sufficiently long waiting period, proceeding during a certain time at a decreasing rate, until it is again stopped. Both the length of the waiting period and the duration and rate of the "delayed" deformation can be assumed to be the result of diffusion processes of particles which, initially blocking this deformation, are first induced by changes in their potential due to the applied forces to migrate and thus to unblock the path of the deformation, but are subsequently drawn back into the region where the deformation proceeds, and precipitate on the path of the motion, as a result of the increasing amount of disorder and, therefore, increasing instability of the structure along this path. Since the rate at which the particles migrate into the distorted structure increases with increasing extent of the distortion, a certain equilibrium position will establish itself when the rate at which the deformation is being retarded by the diffusion of the blocking particles into the deformed regions attains the rate at which it is forced to proceed, thus resulting in a new blocking of the deformation.

The response of metals to loads applied at very high rates, such as impact loads, may also be strongly affected by diffusion processes of foreign particles in the lattice of the material. When the rate at which the force-induced diffusion, unblocking the path of the irrecoverable deformation, proceeds, is higher than the rate at which the elastic deformation increases under the action of the applied forces, the resulting total deformation

will have an appreciable inelastic component. This component can be sharply reduced or even practically eliminated by reducing the diffusion rate (for instance, by lowering the temperature) or by increasing the load rate.

The changes in mechanical properties due to diffusion of foreign particles into previously deformed regions is usually termed *strain aging*. Because of the underlying mechanism of this process, both the time and the temperature dependence of strain aging are closely related to the time and temperature dependence of the diffusion process expressed by eqs. 16·1 and 16·3.<sup>16·3</sup>

### 17. Forced Change of State (Inelastic Deformation)

Changes of the structural pattern produced by the application of external forces remain reversible as long as no particle acquires sufficient energy to migrate from its momentary position of equilibrium to one of lower energy. In the case of a forced change of state the necessary energy is provided by the applied forces in addition to the energy of thermal oscillation of the particles. Thus the changes of structural pattern occurring under the action of the applied forces are the result of the combination of forced and spontaneous changes; therefore, the change of state due to external forces must also be governed by equations of the type of eq. 15·6.

Change of place of an individual particle requires the activation of the particle by the combined action of external forces and thermal energy fluctuations. The activation energy  $q_i$  of the particle in position  $i$  must therefore be provided by the energy of the external forces  $\epsilon_s$  and the energy of the thermal oscillations  $\epsilon_i$ . With  $q_i = \epsilon_s + \epsilon_i$ , the thermal energy required to activate the particle  $\epsilon_i = q_i - \epsilon_s$ . Hence, according to eq. 15·6 the probability that the individual particle will attain this energy as the result of a thermal energy fluctuation can be expressed in the form,

$$p_i = Ce^{-(q_i - \epsilon_s)/kT} \quad (17\cdot1)$$

Equation 17·1 expresses the probability that an individual particle will be displaced from its equilibrium position under the combined action of a force and of the thermal energy. After having been helped by the thermal energy to remove the particle

from its stable position, the acting force will displace it in its direction until the particle drops into a new equilibrium position with an activation energy  $q_j$  higher than  $(\epsilon_i + \epsilon_s)$ . In this case a thermal energy  $\epsilon_j = q_j - \epsilon_s > \epsilon_i$  would be required to continue the displacement of the particle at the rate defined by the probability  $p_i$ . Under a constant force and, therefore, invariable  $\epsilon_s$ , the motion will thus be suddenly slowed down or practically stopped, as the probability  $p_j$  that the particle will attain the energy level  $(q_j - \epsilon_s)$  becomes a fraction of  $p_i$ .

Irreversible deformation of a body under the action of external forces is produced by the permanent displacement of a sufficiently large number of individual particles to produce a per-

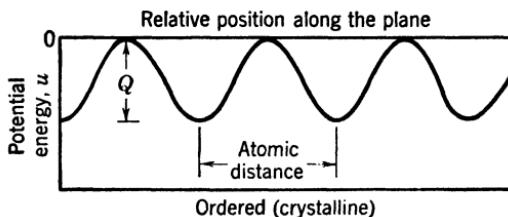


FIG. 17·1 Potential-energy field along an array of ordered particles.

ceptible change of shape of the material body. The displacement of every single particle is governed by equations of the type of eq. 17·1, but the deformation of the material body necessarily depends on the arrangement of the particles within the body. This arrangement determines the field of potential energy of the group of particles and thus their mutual interaction in the course of the deformation.

The potential field in space of an association of particles is built up from the potential fields of each pair of interacting particles. If the array in space of particles is periodic, their field of potential energy will show a periodic pattern; if the particles are arranged at random, the pattern of the potential field will be irregular. Thus, in the ideal crystalline state of matter, potential troughs may be assumed to be regularly spaced and congruent along a lattice plane (Fig. 17·1); along a random section of an amorphous material, on the other hand, the spacing of the troughs, which are of various depth, is irregular (Fig. 17·2). From this rather simplified schematic picture of the difference of the potential energy fields in the ordered and the unordered

state, the fundamental difference between the deformational behavior of crystalline and amorphous materials can be deduced.

Because of the periodicity and congruence of the potential troughs in the ordered crystalline state, the energy ( $q_i - \epsilon_s$ ) necessary to free a particle from its position of equilibrium and to transfer it into an adjacent one will be identical for all particles. Therefore, if a certain amount of energy is applied, particles along one lattice plane may be lifted from their equilibrium positions with regard to the particles of a parallel neighbor-

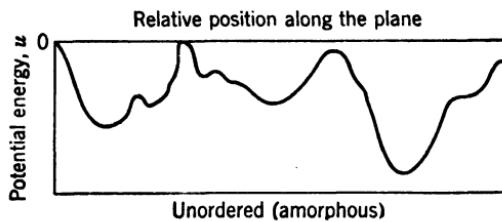


FIG. 17·2 Potential-energy field along an array of particles of random distribution.

ing plane, shifted parallel to this plane and on removal of the force deposited in new equilibrium positions which they fit as closely as the initial ones; this motion along crystal planes is characteristic of the ordered state of matter. Under a constant level of applied energy  $\epsilon_s$ , it goes on until its progress is blocked by imperfections in the lattice or by crystal boundaries; it is called *slip* and is characteristic of the type of deformational behavior usually referred to as *plasticity*.

Although after the termination of a *finished* slip process one layer of atoms has slipped over a parallel layer, the total force required to initiate this slip process is not the sum of the atomic forces resisting the motion along the plane of slip. It is assumed that the slip process starts locally, at a point of lattice imperfection (dislocation) or within a block boundary, by the displacement of an individual atom in the direction of the shear force and that this *glide step* is subsequently propagated through the crystal along the slip plane. Thus the resistance to slip of the atoms within a plane is overcome by overcoming the individual resistances one at a time, with the result that the force required to produce slip is only a fraction of the force that would be necessary to overcome simultaneously the sum of the resistances of all

the atoms involved in the slip process. This assumption concerning the propagation of slip by the "propagation of a dislocation" is supported by the observation of the process of gliding of layers of soap bubbles in Bragg's model of the metallic structure (see Art. 9).

In the course of the slip process the applied energy is dissipated into heat. Below the level of energy  $\epsilon_1$  sufficient to initiate slip within a finite period of time no irrecoverable deformation occurs; particles that have been displaced from their equilibrium position by the applied force have not been shifted over the potential energy barrier (activation energy) between adjacent equilibrium positions. Therefore, on removal of the force, they move back into their initial positions.

The mechanism of slip in real crystals is more complex than the hypothetical mechanism along one slip plane considered in the foregoing discussion. This complexity is due to the fact that slip actually extends over more than the two adjacent lattice planes and is accompanied by rotation of the slip planes, which involves additional distortion of the ordered lattice and that such distortion intensifies the diffusion processes within the crystal. These effects are discussed in the following articles.

In an amorphous substance the energy necessary to lift a particle out of its potential trough varies considerably from particle to particle because of the varying depth of the potential troughs along any one section. If the force provides only a small amount of energy, this may yet be sufficient to move some of the particles out of the shallowest troughs and thus to produce a certain amount of irrecoverable local deformation; the number of particles so affected evidently increases with increasing energy level applied. Since some of the potential troughs are certain to be very shallow, energy is actually dissipated and irrecoverable deformation produced locally by very small forces although at this early stage the process may be so localized as not to lead to perceptible change of shape of the body. Amorphous materials have therefore no definite limit, in terms of the applied force, between the ranges of recoverable and irrecoverable deformation, whereas in crystalline materials this limit, though not rigorously defined, is rather sharply marked.

The probability that any individual particle will leave its position of equilibrium within a group at the bottom of a poten-

tial trough in response to the action of a force depends on the length of time during which this force is applied. As the place change is accompanied by the dissipation through heat radiation of the thermal energy fluctuation, the amount of dissipated energy increases with the period of load application. Therefore the shorter this period, the lower the amount of energy that can be dissipated. Increasing temperatures, by intensifying the thermal oscillations, tend to reduce the waiting time required for a fluctuation of an intensity sufficient to produce activation. Hence, a pronounced influence must be expected of the period of application of the forces and of temperature on the deformational behavior of amorphous materials.

The irregular potential energy pattern of amorphous substances also provides an explanation of their brittleness under rapidly applied forces or at low temperatures. If it were attempted to enforce, at a rate exceeding that of energy dissipation, a certain motion along a random section of the body, the enforced separation between some of the particles in this section, which have no alternative equilibrium position to fit into, would produce atomic bond disruption instead of stabilization of the equilibrium positions of the particles by place change. Such disruption is prevented, however, at higher temperatures or under slowly applied forces, because the particles of an amorphous solid which are always loosely packed, can change place with comparative ease, if the temperature is high enough to supply frequent thermal energy impulses of sufficient intensity. The rate of spontaneous place changes of particles is thus an important factor in determining the extent of the forced change of state of an amorphous substance. Hence, the rate of inelastic deformation of an amorphous material is as temperature-sensitive as the rate of spontaneous place change of particles, that is, the rate of self-diffusion.

The influence of activation by thermal energy in the ordered (crystalline) structure is comparatively small, because the atoms in a crystal are tightly packed; shifting is possible only if it involves a considerable area of the lattice plane. This, however, would require a large thermal energy impulse extending simultaneously or propagating rapidly over the whole area of slip. Since the probability of occurrence of an energy fluctuation of the required intensity simultaneously or consecutively in a very

large number of particles along a potential slip plane is extremely small, deformation by slip will not be substantially facilitated by thermal lattice oscillations. Hence, in nonmetallic (ionic and homopolar) crystals, atomic bond disruption will as a rule occur without appreciable deformation, since the atoms cannot be transferred into new equilibrium positions without permanent disruption of the interatomic bonds unless the thermal activation is extremely high (high temperature, high pressure). In metals, however, the ease with which these bonds are re-established in the course of the deformation (see Art. 9) makes the help of the thermal energy in preventing bond separation much less important. This explains why in metals the influence of the rate of force application is relatively small and why most metals retain appreciable ductility even at very low temperatures.

Since the deformational character of a material is essentially determined by the type of its internal structure, the decisive influence in forming this character in materials in which ordered and unordered phases exist simultaneously is the relative significance of these phases. The most important example of such a material is the polycrystalline metal. In discussing the formation of polycrystalline solid metals out of their melts (see Art. 13), it has been pointed out that, along the boundaries of randomly oriented ordered regions as represented by blocks and crystal grains, some particles will exist in a state of relative disorder because they do not belong to the order of either of the lattices of adjacent domains, being affected simultaneously by the forces exerted by particles in the nearest lattice planes of both domains. Because of their particular state, particles in those boundary regions necessarily produce the deformation effects associated with "viscous" or "relaxing" boundaries;<sup>17·1</sup> for the same reason slip planes will, after formation, show a certain amount of viscosity,<sup>17·2</sup> expressing the disorder introduced into the adjacent lattice structure by the rotation and elastic distortion associated with slip in real crystals (see Art. 18). It should be understood that the "disorder" associated with the arrangement of particles in grain boundaries and slip planes is, in fact, only an intermediate state in the order-disorder spectrum; it can be considered as disorder only in comparison with the almost perfect order existing within the crystal grains.

The viscous or relaxing intercrystalline boundary regions and

slip planes are very thin and, because of their comparative disorder, thermodynamically less stable than the interior of the crystal grains; their response to external forces therefore, is much more sensitive with respect to both temperature and rate of force application than is that of the crystals. The comparative rigidities under the action of forces of the crystalline and the intercrystalline regions will, therefore, change with temperature and rate of loading. At temperatures and loading rates at which the rate of energy dissipation by thermal activation within the intercrystalline boundaries is higher than, or of the same order of magnitude as, the rate of energy produced by the applied force, the boundaries will appear soft in comparison to the crystal areas; however, if the rate of energy dissipation within the boundaries is negligibly small in comparison to the rate of energy application, the intercrystalline spaces will appear rigid in comparison to the crystalline region. This interrelation of loading rate, temperature, and the relative deformational responses of the crystals and intercrystalline regions (boundaries and slip planes) appears to be the most important single factor determining the deformational character and the nature of polycrystalline metals.

The distorted block- and grain-boundary regions and slip planes are the natural locations where foreign particles and impurities tend to aggregate as a result of diffusion processes; this aggregation necessarily affects the deformational behavior of the boundaries as well as their temperature sensitivity. By influencing the thermal stability of the boundary regions foreign particles may be beneficial if the thermal stability is increased (alloying elements), or they may be damaging if the cohesive bonds are weakened by the formation of new molecules (sulfides, oxides).

### 18. Geometrical Aspect of Plastic Deformation

The principal feature of the plastic deformation of crystals is its well-defined geometrical character. Plastic deformation consists of the motion of lamellas of the crystals over each other, along planes parallel to simple crystallographic planes. In this motion the lattice dimensions and the density of the crystals remain practically unchanged. Thus, plastic deformation is concentrated in a succession of parallel slip planes, along

which the motion takes place in certain directions in the planes. The direction of slip has been found to coincide with a crystallographic direction of maximum linear atomic density (that is, maximum number of lattice points per unit length). The slip planes are, in general, parallel to those crystallographic planes containing the slip direction, on which the atomic density (that is, the number of lattice points per unit area) exceeds or is equal to that of any other crystallographic plane containing the same slip direction. Although the rule regarding the slip direction has been observed to hold invariably, the rule of the slip planes is not entirely reliable. In lattices of highest coordination number, that is, in the closest cubic and hexagonal packing, slip directions and planes are determined by this rule. Thus in the face-centered cubic lattice slip occurs on the (111) planes, which are the planes of greatest atomic density, and in the direction [110], which are the directions of greatest atomic density. In the body-centered cubic lattice, however, slip can occur on several planes, all of which are among the most densely packed ones, such as the (110), (112), and (123) planes; the slip direction however, remains invariably the [111] direction. The variation in the slip planes is probably due to the fact that in the body-centered lattice with coordination number 8 the energy content of the bonds of the 6 less near neighbors is of considerable influence in determining the stability of the atomic arrangement within the crystal planes; the energy distribution is, therefore, less definite than in the case of the 12 equivalent nearest neighbors of the face-centered lattice. Several metals such as aluminum and magnesium have different slip planes at low and at high temperature; the slip directions, however, are independent of temperature.

A slip plane and a slip direction in that plane constitute an operative *slip system*. Face-centered cubic crystals with four (111) planes and three slip directions in each plane such as  $\gamma$ -iron thus possess 12 slip systems. Body-centered cubic metals, such as  $\alpha$  iron have 12 equivalent slip planes for each of the four [111] directions, making a total of 48 slip systems.

Among the equivalent slip systems in a crystal the system that actually becomes operative is determined by the *critical resolved shear stress* in the slip plane and in the slip direction.<sup>18-1</sup> If an axial force  $F$  is applied to a crystal of section  $A$ , the resolved

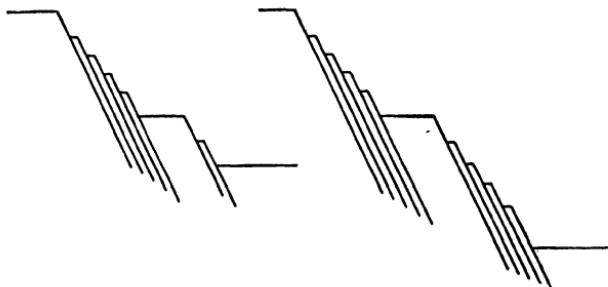
shear stress  $t_s$  is obtained as the force per unit area of the slip plane ( $F/A$ )  $\sin \chi_0 = t_{s0} \sin \chi_0$ , where  $\chi_0$  denotes the angle between force and slip plane, resolved in the slip direction. Hence  $t_s = t_{s0} \sin \chi_0 \cos \lambda_0$ , where  $\lambda_0$  is the angle between force and slip direction.

The critical resolved shear stress at which slip actually starts is different for different crystallographic planes of the same crystal and for different temperatures. Since the initiation of slip depends on the activation energy of the particles in the potential slip plane, which is defined by the depth of the potential troughs between neighboring particles, that is, by the atomic density in the plane, slip along crystallographic planes of different atomic density will necessarily start at a different critical resolved shear stress. Moreover, the critical shear stress differs widely for different materials, as it depends on the magnitude of the interatomic forces. If two planes of equal critical shear stress are subject to the same stress, slip will occur on both planes simultaneously.

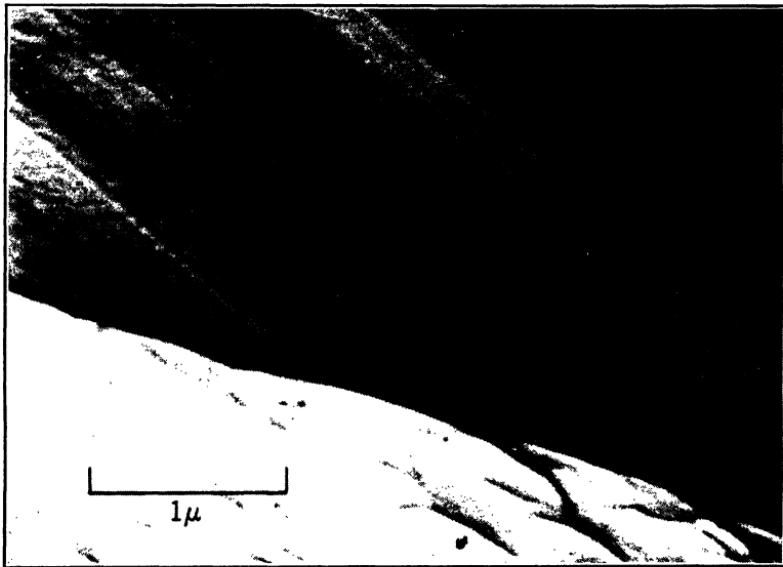
Not all slip planes of any one slip system operate simultaneously; the slip motion is essentially concentrated within a number of favored planes so that *glide laminas* are produced. Studies of the slip process in pure aluminum crystals by the electron microscope and by electron diffraction<sup>18-2</sup> have led to the conclusion that the thickness of individual glide laminas is about 200 Å and that the maximum relative displacement of two adjacent laminas along their common slip plane does not exceed 2000 Å. After such displacement has occurred, a new parallel slip plane becomes operative, producing a similar relative displacement, and so on. Thus the number of laminar slip planes increases with increasing total deformation, as indicated in Fig. 18·1, the slip process resembling that of the shearing of a pack of cards with limited relative displacement of any pair of adjacent cards. The limitation of the relative displacement in laminar slip may be due to the slight rotation of the slip laminas in the deformation process by which the slip resistance in the operating plane is increased, or to the blocking of slip as a result of newly formed dislocation or of intensified diffusion into the regions around such dislocations (see Art. 16).

The operative slip planes are usually not uniformly distributed over wide regions of the deformed body; slip tends to develop in

clusters of closely spaced laminas, with regions of apparently undistorted or only slightly distorted material between them (Fig. 18·1). Although the formation of individual slip laminas



A Schematic illustration of slip band development.



B Electron micrograph of field of laminas in aluminum crystal.

FIG. 18·1 Development of laminar slip and glide lamellas with increasing deformation (after Heidenreich and Shockley,<sup>18·2</sup> courtesy Dr. R. D. Heidenreich).

by relative displacement along one slip plane cannot be optically resolved by any instrument with a smaller resolving power than the electron microscope, the formation of the clusters of laminas is visible, first under the microscope and in an advanced stage

by the eye, as the apparent formation of *slip bands* or *glide lines* (Lueders lines). The smallest distance between these visible glide lines, which is the thickness of the thinnest *glide lamella*, (the apparently undistorted region between glide lines) has been found to be of the order of magnitude of  $10^{-5}$  cm, which is the magnitude of the smallest thermally stable ordered domain (crystallite or block).<sup>18-3</sup>

The relatively high energy content of the atoms in the distorted block boundaries reduces the energy  $\epsilon_s$ , which has to be provided by the external forces before slip is initiated there. Since, according to eq. 17·1, the rate of slip is governed by the exponential of  $(q - \epsilon_s)$ , the effect of a small increase of the potential energy of the particle as expressed by an apparent reduction of  $q_i$  may very considerably increase the probability of the particle being permanently displaced by the action of a force if  $q_i$  and  $\epsilon_s$  are of the same order of magnitude. The slip initiation is, therefore, not governed by the properties of the ideal crystal but by the existing imperfections within the lattice.

Because the operative slip planes are not uniformly distributed but tend to develop in clusters of closely spaced laminar planes, forming slip bands with slightly distorted lamellar regions between them, the appearance of such *slip bands* must be associated with the creation of a certain amount of disorder in the structure, which is the reason for the observed viscous behavior of the material in the slip bands.<sup>17-2</sup>

Visible slip bands are not necessarily made up of slip planes belonging to the same slip system. If, as in the body-centered cubic lattice, the number of nearly equivalent slip systems is large, slip will usually proceed simultaneously or alternately along a number of slip systems. Thus the observable slip bands will not always follow a sharply defined crystallographic plane but may contain slip planes associated with various slip systems.

Since a crystal possesses, in general, several slip systems, different systems may become operative under different conditions, either alternately or simultaneously. It is the number of operative slip systems in a crystal that determines the ease with which plastic deformation proceeds.

In case of slip along different slip systems the final state of deformation is, in general, not independent of the sequence of slip on different planes; the difference in the final shape reached

by different slip sequences is, however, the smaller, the smaller the movements in the individual slip processes along the different slip systems by which the final shape is reached.

It can be shown by a purely geometrical analysis of the possible combinations of slip motions that any small general deformation of a crystal can be produced by slip along five independent slip systems, under the assumption that slip proceeds uniformly along all potential slip planes, that is, without the formation of definite glide lamellas.<sup>18-4</sup> In crystal lattices with less than five independent slip systems a general deformation cannot be produced without rotation, elastic distortion, and bending of the lamellas. Since slip in real crystals is not uniformly distributed over all planes of one slip system, but is associated with the formation of glide lamellas, the general plastic deformation of a real crystal invariably involves elastic distortion of the rotating glide lamellas, by which the simple character of the pure slip process is destroyed even if five independent slip systems exist. However, the appearance of glide lamellas does not alter the fact that crystals with five or more slip systems such as cubic crystals are more easily deformed than crystals with a smaller number of slip systems, for instance hexagonal crystals, which have only one slip plane. In the case of five or more slip planes the distortion and bending of glide lamellas is but a secondary effect, whereas for less than five slip planes the final deformation is the result of the combination of slip and of elastic distortion and bending of the lattice.

In crystals with more than five slip systems, for instance in the cubic lattice, the deformation can take place along different combinations of five independent slip planes. The number of these combinations is very large, although not all possibilities of selecting 5 out of a larger number of slip systems result in 5 systems that are mutually independent; there are, for instance, only two independent slip directions in any one slip plane, whereas all the possible directions are automatically considered in forming the possible combinations of slip systems.

Pure slip without rotation of slip planes can be produced only if the external forces act in the direction of the slip planes; the setting up of such conditions requires an elaborate arrangement of crystal and applied force.<sup>18-5</sup> In the cases of axial tension or compression the slip planes rotate; however, as they are not bent,

these are conditions of practically pure slip, unless the glide lamellae become exceptionally thick.

When the active slip planes rotate away from the direction of the acting force, as in axial compression (Fig. 18·2), the resolved

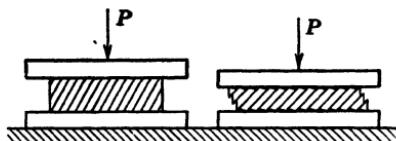


FIG. 18·2 Compression of crystal.

shear stress on the slip planes diminishes; at the same time the resolved shear stress on a potential alternative slip system increases, until the slip motion is transferred from the first slip system to the second. After rotation of the second plane, the reverse transfer takes place, so that slip proceeds alternately on both planes. In tension, on the other hand, the operative slip planes rotate into the direction of the acting force (Fig. 18·3), and slip must be expected to proceed along the initial planes, resulting finally in the formation of preferred orientation in the direction of the force. This relation between the direction of the force and the formation of slip planes leads to different deformational behavior of the single crystal in tension and in compression.

Translatory slip is not the only possible type of plastic deformation in which no change in the periodicity and symmetry of the lattice takes place. Under certain conditions a particular slip



FIG. 18·3 Extension of crystal.

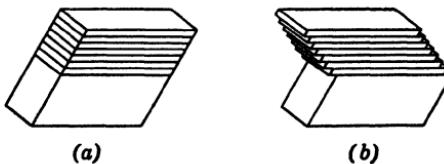


FIG. 18·4 Schematic representation of twinning. (a) → (b).

plane may become a mirror plane, one part of the crystal having become in the course of the deformation the mirror image of the other part from which it is separated by a sharp line (Fig. 18·4). This process, which is called *twinning*, is usually rather abrupt and frequently accompanied by a sharp sound ("crying" of tin in bending). In plastic deformation the direct influence of twinning, in general, is relatively insignificant; it may, how-

ever, become a major indirect factor initiating extensive slip, particularly in crystals with less than five slip systems (such as zinc and tin), where the operation of a twinning plane suddenly produces favorable conditions for the operation of new slip systems which were unfavorably oriented before the twinning occurred. The movement of atoms in the lattice, in general, is identical in both translatory slip and in twinning. The critical resolved shear at which twinning starts, however, has been found to be higher than the critical resolved shear stress.<sup>18-6</sup>

Under certain conditions slip can lead to a phase transformation, the most important example of such a process being the *martensitic transformation* of steel. This is a transformation from the face-centered cubic (austenitic)  $\gamma$  structure to the body-centered tetragonal (martensitic) structure produced by slip along one of the (111) planes of the  $\gamma$  phase, in which the carbon atoms occupying the centers of the cube do not move and, therefore, block the slip process before the body-centered cubic (ferritic)  $\alpha$  structure is reached.<sup>18-7</sup> This blocking effect of the carbon atoms is achieved, however, only if the cooling rate accompanying the deformation is so high that the carbon atoms cannot diffuse out of the transformed lattice. The prematurely arrested slip produces the characteristic tetragonal lattice of martensite which appears in the form of sharp "needles" representing shapes of minimum energy with respect to the textural stresses set up by the transformation both in the martensite and in the adjacent untransformed austenite, by which the martensitic hardening effect of the steel is caused. Similar transformations by arrested slip occur in other metals; they are all based on similar mechanisms.

## 19. Structural Theory of Inelastic Deformation

Attempts to derive equations describing inelastic deformation in terms of atomic structure are based on the assumption that the energy distribution over the atoms can be expressed by the Boltzmann-Maxwell-Gibbs distribution function. Thus the probability of occurrence within the atomic structure of spontaneous changes of thermal energy which initiate place changes of atoms is given by eq. 15-7 in which  $\epsilon = q$ .

In the absence of external forces the probability of spontaneous place changes of particles, or the number of changes per second

expressed by  $\nu_0 p$ , will necessarily be the same in all directions, as there is no reason for any directional preference. However, if an external force is applied in one direction, the number of place changes or the *rate of activation* in this direction will not remain the same as before, since the energy level a particle must attain momentarily in order to be able to move in the direction of the acting force across the potential barrier into an alternative position of equilibrium has been reduced by the potential energy  $\epsilon_s$  of the applied force, whereas it has been increased by the same amount with respect to the possible movement in the opposite direction. The momentary energy level required for activation in the direction of the applied force has thus been reduced to  $(q - \epsilon_s)$ , whereas it has been increased with respect to activation in the opposite direction to  $(q + \epsilon_s)$ . Hence, the rate of activation in the direction of the applied force for one mole of the substance is given by

$$u_1 = \nu_0 C e^{-(q - \epsilon_s)/RT} = \nu_0 C e^{-Q/RT} \cdot e^{\epsilon_s/RT} \quad (19 \cdot 1)$$

in the opposite direction,

$$u_2 = \nu_0 C e^{-(q + \epsilon_s)/RT} = \nu_0 C e^{-Q/RT} \cdot e^{-\epsilon_s/RT} \quad (19 \cdot 2)$$

if  $\epsilon_s$  is assumed to denote the energy of the external force per mole. Hence, the resulting rate of deformation in the direction of the applied force,

$$\begin{aligned} u &= u_1 - u_2 = \nu_0 C e^{-Q/RT} \cdot (e^{\epsilon_s/RT} - e^{-\epsilon_s/RT}) \\ &= \text{const } e^{-Q/RT} \sinh (\epsilon_s/RT) \end{aligned} \quad (19 \cdot 3)$$

At temperatures below the Debye temperature this equation takes the form,

$$u = \text{const } \frac{kT}{h} \cdot e^{-Q/RT} \sinh (\epsilon_s/RT) \quad (19 \cdot 4)$$

Equations 19·3 and 19·4 will now be considered under two limiting conditions:

1. If the potential energy of the applied force  $\epsilon_s$  is relatively small in comparison with the thermal energy  $RT$  so that  $(\epsilon_s/RT)$  is small, the function  $\sinh (\epsilon_s/RT)$  can be approximated by  $(\epsilon_s/RT)$ . Hence, the velocity of deformation, according to eq. 19·3,

$$u = \text{const } \epsilon_s e^{-Q/RT} \quad (19 \cdot 5)$$

If eq. 19·5 is written in the form,

$$\epsilon_s = \text{const } ue^{Q/RT} = \eta u \quad (19\cdot6)$$

the coefficient of viscosity  $\eta = \text{const } e^{Q/RT}$  expresses the resistance to deformation. Hence a straight-line relation should exist,  $\log \eta = \text{const } 1/T$ ; the existence of such a relation has been verified by experiment. It is evident that the stress required to produce inelastic deformation at a given elastic strain  $e_0$  depends on the velocity of deformation and on temperature only through the combined velocity-temperature parameter ( $\text{const } ue^{Q/RT}$ ), and not on velocity of deformation and temperature independently. The relation between the velocity of deformation and the temperature producing a given stress  $s$  at a given elastic strain is, therefore,

$$ue^{Q/RT} = \text{const } s \quad (19\cdot7)$$

The expression,

$$T(\text{const} - \log u) = T_m \quad (19\cdot8)$$

defines a velocity-modified temperature  $T_m^{19\cdot1}$  on which depends the stress response  $s$  in mechanical tests performed at certain constant strain rates, or  $s = f(T_m, e_0)$ .

Equations 19·5 to 19·8 describe the type of temperature and velocity-sensitive deformational behavior that is usually designated as *viscosity, relaxation* or, more generally, *thermal inelasticity*.

2. If the energy of the applied force  $\epsilon_s$  is much larger than the thermal energy  $RT$  facilitating the motion, the function  $\sinh(\epsilon_s/RT)$  may be replaced by  $\frac{1}{2}e^{\epsilon_s/RT}$ , and eq. 19·4 be written in the form,

$$u = \text{const } \frac{kT}{h} e^{-(Q-\epsilon_s)/RT} = \text{const } \frac{kT}{h} e^{-Q/RT} \cdot e^{\epsilon_s/RT} \quad (19\cdot9)$$

which defines the so-called *plasticity* or *athermal inelasticity* in which the influence of thermal activation is very small and most of the energy necessary to produce the deformation must be supplied by the external force.

The presented derivation of eqs. 19·5 and 19·9 shows that *thermal inelasticity* and *athermal inelasticity* represent both limiting cases of general inelasticity governed by eqs. 19·3 and 19·4. Under conditions for which the potential energy of the applied forces and the thermal-energy fluctuations within the

structure are of comparable magnitude, the deformational behavior must be described by the hyperbolic sine relation of general inelasticity. Equation 19·5 shows that, as long as the applied energy  $\epsilon_s$  is small or the temperature high, the rate of deformation increases directly with  $\epsilon_s$ ; on the other hand, when the applied energy is relatively large or the temperature low, eq. 19·9 expresses the fact that a small change in the applied energy  $\epsilon_s$  produces a considerable change in the velocity of the deformation.

Theoretically eq. 19·4 gives a finite value of the velocity  $u$  for all values of  $\epsilon_s$ ; however, this velocity will be observable only if the resulting deformation is within the range of observation of the optical instruments used. The deformation within finite time becomes imperceptible when the energy  $\epsilon_s$  of the applied force drops below a critical limit. This energy limit, which represents the potential energy associated with the critical resolved shear stress at which slip starts, is called *resilience*; it cannot be defined in absolute terms, however, but only in terms of a value of the velocity of deformation  $u$  designated as *observable*. Since only at absolute zero temperature or at zero force could the velocity of deformation become zero, these are the two limiting conditions for which an absolute elastic limit defining the ranges of reversible and irreversible deformation could possibly exist. Hence, for any temperature  $T > 0$  and loading of relatively long duration, a theoretical elastic limit could only be associated with zero stress. For short loading periods, however, the designated or *conventional* elastic limit can be defined in terms of the critical shear stress  $s_0$  or of the resilience  $\epsilon_{s0}$  producing an arbitrarily small *measurable* permanent deformation; the pertaining velocity of deformation  $u_0$  is obtained from eq. 19·9 in the general form,

$$\epsilon_{s0} = Q - RT [\text{const} - \log (u_0/T)] \quad (19\cdot10)$$

Thus, the stress  $s_0$  at a definite elastic strain  $e_0$  is again a function of a velocity-modified temperature  $T[1 - \text{const} \log (u/T)] = T_m$ . The general relations between resilience and velocity of deformation at constant temperature and between resilience and temperature at constant velocity of deformation are obtained from eq. 19·10, in the form,

$$\epsilon_{s0} = C_1 + C_2 \log u_0 \quad (19\cdot11)$$

and

$$\epsilon_{s0} \doteq C_3 - C_4 T \quad (19 \cdot 12)$$

Since the resilience  $\epsilon_{s0}$  is proportional to the square of the critical shear stress, the relations between the yield limit  $s_0$  (critical shear stress) and either velocity of deformation at constant temperature or temperature at constant velocity of deformation can be written in the form,

$$s_0 = c_1 + c_2 \sqrt{\log u_0} \quad (19 \cdot 13)$$

and

$$s_0 = c_3 - c_4 \sqrt{T} \quad (19 \cdot 14)$$

The validity of these two relations has been confirmed by the experimental results obtained for a number of metal crystals.<sup>19·2</sup>

If the process of *continuing* inelastic deformation itself is considered, not only its *initiation*, the applied shear stress must always act through a distance equal to one-half the interatomic distance  $d$  before any particle is brought into the intermediate state of high energy at which the "spontaneous" movement of the particle into the adjacent potential trough can take place. In order to repeat this process continually and thus to produce deformation at a certain velocity  $u$ , the energy per mole of the applied stress  $s : \epsilon_s = \frac{1}{2}Nds = \alpha s$ . If this expression is introduced into eq. 19·3 the general equation of flow is obtained,

$$u = \text{const } e^{-Q/RT} \sinh (\alpha s/RT) \quad (19 \cdot 15)$$

where  $\alpha$  is a constant which depends on the atomic or molecular structure. At constant temperature,

$$u = C \sinh cs \quad (19 \cdot 16)$$

where both constants  $C$  and  $c$  depend on temperature and on structure.

For small stresses eq. 19·16 is transformed into the Newtonian relation of linear viscous flow,

$$u = \text{const } s \quad (19 \cdot 17)$$

for large stresses into the well-known logarithmic relation,

$$u = \text{const } e^{cs} \quad \text{or} \quad \log u = \text{const} + cs \quad (19 \cdot 18)$$

A hyperbolic sine relation between  $u$  and  $s$  has been derived

from a model of a dislocation by Prandtl<sup>19·3</sup> and suggested on a purely empirical basis by Nadai.<sup>19·4</sup> It has recently been developed from the theory of rate process by various investigators.<sup>19·5</sup>

## 20. Complex Types of Deformation

The simple mechanisms of deformation—elasticity, viscosity (relaxation), and slip—are limiting types associated with idealized materials. The deformation of real materials is in general produced by all three mechanisms of simple deformation, combined with the complex mechanisms that are brought into play by the changes of internal structure associated with large inelastic deformation, and that are modified in the course of this deformation by the temperature-induced processes of diffusion and thermal changes in the structure; these processes, going on simultaneously with the deformation, are themselves modified as a result of the preceding deformation.

**TEXTURE FORMATION. CHANGE IN MOLECULAR STRUCTURE.** Force-induced changes of the internal structure in metals may be the result of extensive slip by which the dimensions of the undistorted crystal domains are reduced (*crystal fragmentation*), considerable disorder is introduced into the structure as a result of bending and distortion of fragments, and the resistance to further deformation is increased (*strain hardening*). With increasing deformation the fragmentation process gradually changes into a process of development of preferred orientation of the crystal fragments in the direction of the maximum velocity of deformation (*texture formation*); such process, however, would be impossible without the help of thermal activation and thus constitutes in fact a combination of fragmentation, recrystallization, and the development of texture. Even the fragmentation process itself is associated with simultaneous thermal changes in the fragmented structure: since slip planes are about 200 Å apart, slip along two or three operative slip systems would produce laminar rods or laminar blocks of 200 Å side length; rods or blocks of such dimension, however, are thermally unstable as the high interfacial energy produces rapid recrystallization of those blocks to the minimum stable size of ordered domains at the given temperature. Thermal changes which cannot be optically resolved are usually termed *recovery*.

Force-induced changes of internal structure in molecular sub-

stances are readily identifiable only in *macromolecular* materials, such as polymers; in the *micromolecular monomers*, which are substances built up of small molecules, such as glass, the changes are of a statistical nature producing the simple deformation mechanism of viscosity. Changes of internal structure after considerable deformation in polymers, particularly in the weakly cross-linked rubber-like substances, fibers, and filaments may lead to the development of a preferred orientation by stretching of the chain molecules in the direction of the maximum velocity of deformation. Because of the ordered arrangement of the stretched parallel chain molecules the process by which this structure is produced is termed "*crystallization*."<sup>20-1</sup> Since the unstretched molecules of materials in which crystallization may be produced by large deformation are assumed to have the shape of coiled springs and are, in this shape, interlinked by weak secondary bonds, the stretching process must be preceded by the loosening or disruption under the action of the applied forces of those molecular bonds which prevent the molecules from uncoiling. The *change of configuration* which is initiated by the disruption of the intermolecular bonds may start abruptly if all or most of the bonds are broken simultaneously; it is then accompanied by the same abrupt change of the deformational response to the applied forces that is characteristic of the sharp yield point observed in metal crystals and certain polycrystals; in other materials the change is more gradual, indicating a successive loosening and breakdown of the molecular bonds. In both cases the material appears to "*soften*" during the change of configuration, until the effects of the orientation of the molecules in the direction of the deformation produce an apparent *stiffening* or hardening of the substance, as indicated by the characteristic shape of the stress-strain diagrams of rubber and textile fibers (Fig. 20·1). In certain high polymers, for instance, in polystyrenes and methacrylates (Lucite, Plexiglas), the apparent softening at low strain rates proceeds until fracture occurs after substantial deformation at constant or even decreasing stress, without any "*stiffening*." In this case the bond disruptions probably extend to interatomic bonds, and the deformation proceeding at a small strain rate and at decreasing or constant stress represents in fact a progressive fracture process. This conclusion is borne out by the observation, at this stage, of a

net of very fine cracks covering the most highly stressed parts of the surface ("crazing").

The apparent softening of molecular substances produced by the breakdown of secondary bonds under the action of forces is usually known as *thixotropy*; after the forces have been released new molecular bonds establish themselves in time as a result of thermal activation and the reaction between the molecules; the weakly bound molecular structure thus tends to revert to its initial configuration.

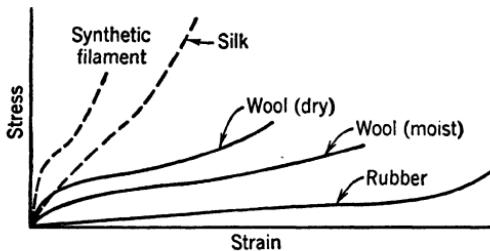


FIG. 20·1 Schematic stress-strain diagrams of various fibers.

*Crystallization* of high polymers constitutes a permanent change of structure and thus a change of state only if the stretching force is sustained long enough for new cross links to be formed between the stretched macromolecules before the force is removed, stabilizing their new shape. The ordered configuration of the crystallized structure will necessarily be the more stable, the lower the temperature. The intensified molecular diffusion and chemical reactions taking place at elevated temperatures, by dissolving the cross links that stabilize the crystallized structure, cause necessarily a more or less rapid loss of order and thus a *randomization* of the anisotropic structure. Under certain conditions and at certain temperatures the thermal activation and the resulting chemical reactions become strong enough to produce a breakup of the long molecular chains into smaller segments (*depolymerization*), thus extending the randomization process from the molecular network to the individual macromolecules.

**STRAIN HARDENING.** Texture formation in crystalline substances is invariably preceded by plastic slip, crystal fragmentation, and distortion of the fragmented structure. Since slip is not uniformly distributed over all possible slip planes but concentrated within a limited number of planes and therefore pro-

duces glide lamallas of finite thickness, plastic deformation of a crystal is always associated with rotation, bending, and elastic distortion of the glide lamallas, even in the case of the existence of 5 or more slip systems. Thus, even in the single crystal the velocity of deformation  $u$  cannot be sustained by the same force or stress at which it is initiated but requires a gradually increasing stress. When the force is removed and subsequently reapplied, slip at the previous velocity will start at the stress at which the test was interrupted if no load-independent changes occur in the structure. This phenomenon is known as *strain hardening*, *work hardening*, or *cold working*; it is measured by the difference  $\Delta\epsilon_s$  between the resilience, or the difference  $\Delta s$  between the critical shear stress of the crystal in the hardened and in the initial state, associated with the same velocity of deformation  $u$ . The amount of strain hardening is related to the deviation of the real deformation of the crystal from ideal conditions of pure slip proceeding simultaneously or alternately on 5 slip systems. Strain hardening, therefore, will be the more pronounced, the smaller the number of existing slip systems; the less uniform the distribution of slip, that is, the thicker the glide lamellas; and the less homogeneous the field of the forces producing the deformation.

Retardation of slip by work hardening, however, will considerably affect even the progress of deformation under homogeneous stress, proceeding by slip on alternately operating slip planes. Since slip on any one slip system is blocked by work hardening, unless the hardening effect is compensated by the increase of the resolved shear stress resulting from the rotation of the slip planes into the direction of the acting force, slip on alternating planes will usually take place both in tension and in compression. Because of this alternation the development of preferred orientation in tension will be delayed, and the difference between the inhomogeneous plastic deformation in tension and in compression, which is considerable in metals without work hardening, will be partly eliminated.

The increased resilience  $\Delta\epsilon_s$  is the excess energy in the slip planes, associated with the elastic distortion of the glide lamellas. Thus, for instance, in the bending of a crystal the critical shear stress at which slip takes place along the slip planes forming the glide lamallas (Fig. 20·2) is made up of the critical shear stress of the undistorted planes and the additional stress necessary to

produce the distorted shape of the lamellas and to sustain the slip within the bent slip planes. The distortion produces, moreover, an additional effect on the atomic arrangement within the slip planes, which is illustrated in Fig. 20·3. The different

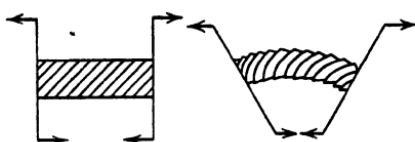


FIG. 20·2 Bending of crystal.

density of the atomic lattices of bent lamellas on both sides of a slip plane (extension on the convex side, compression on the concave) causes displacement relative to each other

of atoms on either side of the slip planes and thus creates dislocations in the planes. These dislocations produce the *atomic* hardening effect which, however, is usually small in comparison with the hardening due to elastic bending and distortion of glide lamellas. In polycrystalline metals the hardening effects due to slip retardation on bent glide planes and to distortion and rotation of glide lamellas within an individual crystal are probably less important than the effects of slip interference at and near grain boundaries. Moreover, the effect of the fragmentation of the crystal structure

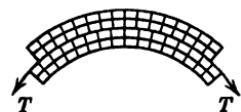


FIG. 20·3 Atomic dislocation along glide plane in bending.

by extensive slip introduces into this structure a system of highly inhomogeneous textural stresses on the microscopic and atomic scale (dislocations), the potential of which is an indication of the hardening of the material. In order to conform easily with the deformations of neighboring grains, crystals making up the aggregate should have at least five and, preferably, more slip systems. Therefore, in an aggregate of crystals which have less than five slip systems, such as hexagonal crystals, slip will, in general, start at a stress many times higher than that producing slip in a single crystal of average orientation, which itself is higher than the critical shear stress on the slip plane. Hence, the difference in the hardening effect between the polycrystalline aggregate and the single crystal is much more pronounced in hexagonal crystals, such as zinc, than in cubic crystals with 12 or 48 slip systems, such as iron and aluminum (Fig. 20·4).

The amount of strain hardening, which is produced by a certain stress, is evidently not independent of time nor of the temperature

during the application as well as after the removal of the force. A deformed single crystal is actually no longer a *single* crystal, as it has lost its character of a single crystal after the first slip has introduced distorted slip planes. Since the energy content of the distorted lattice is higher than that of the undisturbed lattice, it is thermodynamically unstable. Because of the

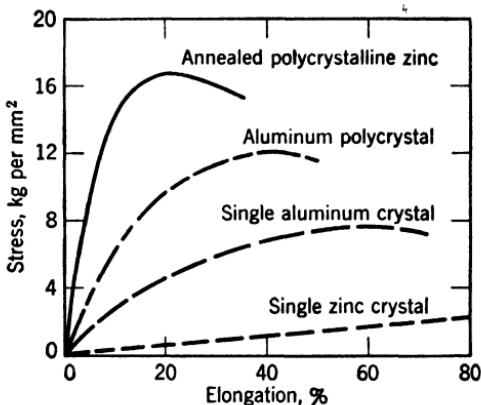


FIG. 20·4 Comparison of stress-strain curves for single crystal and polycrystalline specimen of zinc (after Elam<sup>20·2</sup>) and aluminum (after Karnop and Sachs<sup>20·3</sup>).

"spontaneous" changes going on within the structure at any temperature above absolute zero, the excess energy of the particles associated with previous slip and strain hardening is gradually dissipated by place change of particles in unstable positions. Hence the stress or energy necessary to produce and sustain further plastic slip decreases: the hardening decreases as the initial resilience  $\epsilon_{s0}$  or critical shear stress  $s_0$  is gradually restored or *recovered*. Since the rate of thermal activation depends on temperature according to eq. 16·1 the extent of *recovery* will increase with time and with increasing temperature.

The rate of recovery is proportional to the number  $v$  of place changes per second and to the momentary number of particles existing in states of energy above the stable state, to which they have been transferred in the course of the hardening process. If  $n_0$  denotes the total number of such particles and  $n$  the number of particles that have already been restored to a stable state of energy in the course of the recovery process, the rate ( $dn/dt$ ) at which particles recover their stable condition is proportional to

$\nu(n_0 - n)$ ; hence, with  $\nu = \nu_0 e^{-Q/RT}$  where  $\nu_0$  is given by eq. 16·2:

$$\frac{dn}{dt} = (n_0 - n) \text{ const } \nu_0 e^{-Q/RT} \quad (20 \cdot 1)$$

or .

$$\frac{dn}{(n_0 - n)} = \text{const } \nu_0 e^{-Q/RT} dt \quad (20 \cdot 2)$$

By integration,

$$(n_0 - n) = n_0 e^{-\text{const } \nu_0 e^{-Q/RT} t} \quad (20 \cdot 3)$$

or

$$\frac{n}{n_0} = 1 - e^{-\text{const } \frac{kT}{h} e^{-Q/RT} t} \quad (20 \cdot 4)$$

Equation 20·4 indicates the extent of recovery in terms of the ratio  $n/n_0$  of particles restored to a stable energy level and the

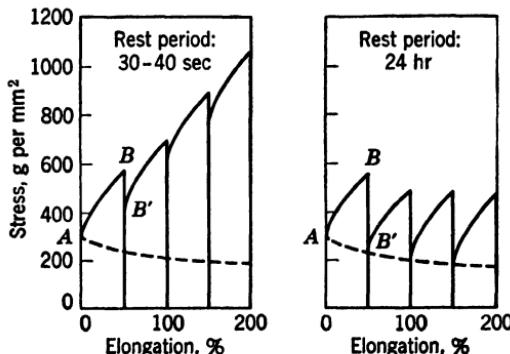


FIG. 20·5 Recovery of zinc crystals (after Haase and Schmid<sup>20·4</sup>).

dependency of recovery on time and temperature. Although theoretically complete recovery with  $n = n_0$  requires infinite time, the times required for extensive recovery to take place in metal crystals at room temperature are relatively short; they can be considerably shortened by the application of elevated temperatures. For zinc crystals for instance, complete recovery at room temperature was observed after 24 hr (Fig. 20·5).

Obviously, recovery takes place not only after but also during the hardening process itself; therefore, the observed hardening is actually the resultant of the true hardening effect as well as of the other diffusion processes going on simultaneously, including

simultaneous recovery. The gradual decrease with increasing deformation of the rate of hardening which can be observed in all metals is, to a certain extent at least, due to the intensified recovery associated with increasing deformation by which the thermal stability of the structure is reduced.

Equation 19·9 which governs the velocity of slip has been derived on the assumption that in the initiation of the slip process the influence of the energy of the applied forces is large compared to that of the heat energy, since the probability of thermal energy fluctuations occurring simultaneously over all particles involved in the slip motion is very small; the effect of temperature on the critical shear stress is, therefore, also small. However, the progress of slip and of work hardening depends very much on temperature. This temperature influence results from the recovery and other diffusion processes going on simultaneously with the deformation. Thus, the velocity of deformation which, according to eq. 19·18, affects the critical shear stress but little, is of considerable influence on the progress of strain hardening, since at a high velocity of deformation the hardening rate must necessarily be higher than at a low velocity at which the softening effect due to recovery is more pronounced.

From eq. 20·4 an equivalence relation can be obtained of the effects of time and of temperature in the recovery process of metal crystals. If it is assumed, to a first approximation, that the true hardening does not itself depend on either velocity of deformation or temperature, this dependency is introduced indirectly by the recovery effect and diffusion processes. The same apparent hardening will, therefore, be associated with an identical extent of recovery if diffusion processes lending to precipitation are insignificant. Hence, the relation between  $t$  and  $T$  producing the same amount of recovery is obtained from eq. 20·4 by introducing  $n/n_0 = \text{const}$ .

$$\nu_0 t e^{-Q/RT} = \text{const} \quad (20\cdot 5)$$

or

$$\log (Tt) = Q/RT + \text{const} \quad (20\cdot 6)$$

valid for  $T < \theta$ . For  $T > \theta$  the left side of eq. 20·6 becomes  $\log t$ . Equation 20·6 shows that the work-hardening process can be considerably more influenced by a change of temperature than

by a variation of the strain rate, that is, a variation of the time  $t$  during which the forces are applied.

Slip in single crystals can be much more rapidly blocked by diffusion of foreign atoms into the operative slip planes than by atomic strain hardening which results from the textural microstresses introduced in the course of the bending and distortion of glide lamellas. By such diffusion the stress limit at which substantial plastic deformation occurs is considerably raised and the apparently continuous, quasihomogeneous hardening process transformed into an abrupt, inhomogeneous yield which, frequently, is initiated at a higher stress than that at which it continues (see Art. 16). The effective blocking of operative slip planes which produces this phenomenon can, however, take place only in crystals with one or two slip planes, as the blocking of one or even of a few slip planes would not produce any appreciable effect on the resulting yield process in crystals in which slip proceeds on a large number of alternative planes.

**CREEP.** Not only are changes in the internal structure of materials by slip, fragmentation, and subsequent formation of texture produced under conditions in which the increasing resistance to deformation due to strain hardening is overcome by increasing stresses, so that a constant rate of deformation is sustained; they are also produced if the stresses are increased, independently of the increase in resistance to deformation, at a lower rate than this resistance increases, even if, in the limiting case, the rate of stress increase is zero; that is, the stress is sustained at a constant level. This follows from the general eq. 19·4, from which a finite value of inelastic deformation can be obtained for any value of  $\epsilon_*$  as long as  $T > 0$ , if only the stress is applied during a sufficiently long period. The deformational response of a material to a sustained stress is designated as *creep*, which is therefore not a particular type of inelasticity, but only the response to a particular type of loading. Therefore, it will be produced by the same combination of simple deformation mechanisms, complex mechanisms of stress-induced structural change, and effects of thermal processes within the structure, which is responsible for inelastic deformation under any other loading condition. The particular importance of creep in engineering and in research, however, is due to the fact that conditions of sustained stress are very frequently encountered in

practice, and that, moreover, tests of materials under conditions of sustained or very slowly increasing stress bring out basic features of deformational response that are obliterated in tests in which relatively high constant rates of deformation are induced. Because of the strain hardening produced in metals by the deformation, the rate of deformation in creep tests of metals decreases with time as long as the hardening proceeds. This follows immediately from a consideration of eq. 19·4 or of the simplified eqs. 19·5 and 19·9, if it is considered that  $\epsilon_s = \text{const } s^2$  and that for a strain-hardening material  $s = f(e)$ , where  $f(e)$  represents a monotonously increasing function of the total strain  $e$ . With  $u = de/dt$  differential equations for  $e$  are obtained from any one of the afore-mentioned equations. Their solutions for arbitrary monotonously increasing strain-hardening functions  $f(e)$  produce creep diagrams  $e(t)$ , the time derivatives of which decrease with time, in accordance with the shape of observed creep curves of metals (Fig. 49·2). This conclusion holds for the creep of single crystals as well as for polycrystalline aggregates. In the latter, however, the effect of the interaction between crystals and intercristalline boundaries will influence the creep behavior considerably.

It can be inferred from eq. 19·15 that the rate of deformation under a constant force is extremely sensitive to changes of temperature. For  $s = \text{const}$ , eq. 19·15 may be written in the form:

$$\log u = C_5 - C_6 \frac{1}{T} \quad (20\cdot7)$$

Because of this relation a change of a few degrees in  $T$  will produce a considerable change in  $u$ . This fact has been confirmed by numerous test results, such as those on zinc and tin crystals indicated in Fig. 20·6, which show that a change of temperature by  $20^\circ\text{C}$  may produce a change by a factor of nearly 10 in the velocity of the deformation.

Creep tests and tests at very low stress rates have been used to demonstrate and investigate the characteristic inhomogeneity of plastic slip in metals, both single crystals and polycrystalline aggregates, which is caused by the precipitation of alloying elements, such as carbon in steel, copper and magnesium in aluminum, beryllium in copper, on the slip planes in the course of the deformation (see Art. 16). These tests<sup>20·6</sup> show alternate

periods of very rapid and very slow creep, producing the step-like stress-strain diagrams characteristic of tests performed at low loading rates (Fig. 20·7). The duration of the strain steps varies with the loading rate and may attain several days and

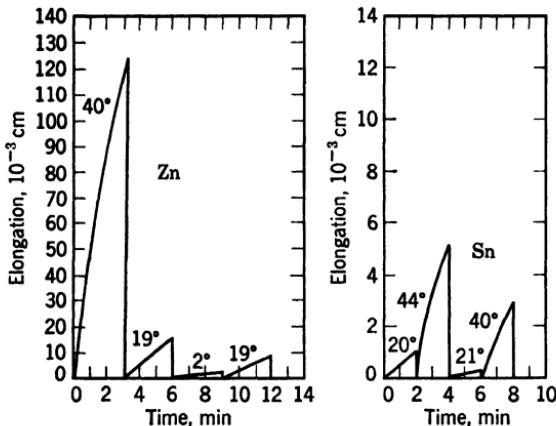


FIG. 20·6 Velocity of slip as function of temperature (after Boas<sup>20·5</sup>).

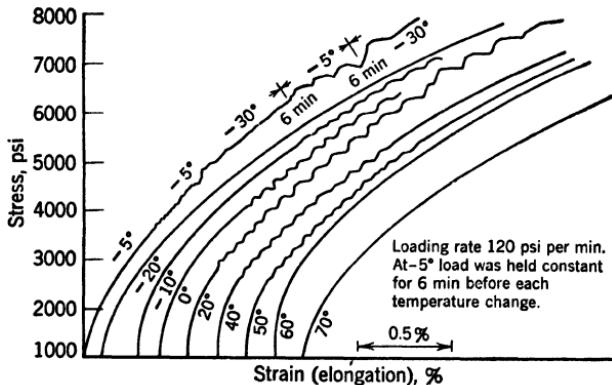


FIG. 20·7 Discontinuous stress-strain curves at various temperatures in aluminum (after McReynolds<sup>20·6</sup>).

even months if the loading rates are extremely low.<sup>20·7</sup> It has been noted that discontinuity of the stress-strain diagrams is characteristic for a certain range of temperatures only; beyond this range, which is bounded by the temperature at which the diffusion of the alloying element becomes too slow for precipitation to be significant, and the temperature at which the creep

rate becomes too high for the precipitation to have a noticeable effect, the stress-strain diagrams appear perfectly continuous. This fact tends to support the assumption that diffusion processes are responsible for the observed inhomogeneity and discontinuity of plastic slip and illustrates the importance of those processes in the interpretation of the deformational response of materials to applied forces.

**CLASSIFICATION OF MECHANISMS OF CHANGE OF STATE.** Because of this importance, the observed inelastic response of a material can be interpreted correctly, and the significance of the observations evaluated, only if the effects of all mechanisms, both force-induced and temperature-(time-)induced, expected to produce a change of state, are considered, and the relative significance of each of these mechanisms is analyzed under the particular conditions of the test or of the performance of the material.

In order to facilitate such analysis, a classification of the operative mechanisms has been proposed in Table 20·1. In this classification the mechanisms are divided into two large groups: one group of mechanisms operative primarily in metals; the other operative primarily in molecular substances, particularly high polymers. Although such division appears reasonable from a general point of view, it is not rigorous. Thus, for instance, viscous deformation, although primarily associated with molecular substances, will not be entirely absent in metals. Similarly, the division into force-induced and temperature- or time-induced mechanisms does not imply that the latter mechanisms are independent of the applied force; the division only intends to convey the fact that they are not directly induced by the applied forces, being affected by them only indirectly by way of the deformation.

Although in every deformation process all the mechanisms associated with the particular type of material are theoretically operative, only certain of those mechanisms will be significant under certain conditions or, for the same conditions, during different periods of the test or the performance. Thus, for instance, in a polycrystalline metal viscosity (relaxation) will dominate the initial creep behavior, if the structure is fine grained, whereas slip will be dominant if the metal is coarse grained. Since viscosity is much more time- and temperature-sensitive than slip, it must be expected that the creep behavior of the

fine-grained metal will be more sensitive to temperature changes than that of the coarse-grained metal. It must further be expected from a consideration of Table 20·1 that, the longer the duration of the creep and the larger the strains, the more significant the effects of the complex mechanisms, such as fragmentation and recrystallization.

TABLE 20·1

## CLASSIFICATION OF MECHANISMS RESPONSIBLE FOR CHANGES OF STATE IN ENGINEERING MATERIALS

		Mechanisms of Change of State	
		Simple	Complex
Involving	Individual particles	Macromolecules and crystals	General character of structure
Force-induced (deformation)			
Metals	Slip	Fragmentation (hardening)	Formation of anisotropic texture
Polymers	Relaxation (viscosity)	Change of configuration of molecular structure (thixotropy)	"Crystallization"
	Small deformations	→ Direction of increasing deformation	
Temperature-(time-)induced			
Metals	Self-diffusion (recovery) precipitation (aging)	Recrystallization	Phase change
Polymers	Molecular diffusion	Chemical reaction	Depolymerization (randomization of structure)
	Moderate temperatures	→ Direction of increasing temperature and time	

The force-induced and the time-induced mechanisms in the individual subgroups are interrelated in such a manner that an

intensification of the former may produce an intensification of the latter, and vice versa. Thus, for instance, intensified fragmentation may be both the cause and the result of intensified recrystallization. On the other hand, the trend may be the reverse, as in the case of intensified slip producing intensified precipitation whereas intensified precipitation retards the slip. In molecular structures the change of configuration produced by the applied force may affect the intensity of certain chemical reactions, such as cross linking, and thus lead to a complete reorganization of the deformed structure, which depends on the intensity of the applied forces but, at the same time, changes the deformational response of the material to those forces. This mutual interrelation of the force-induced and the temperature-induced mechanisms makes the interpretation of an observed deformational response of any material beyond the range of small deformations rather difficult.

## 21. Volumetric Deformation

Inelastic deformation of a homogeneous isotropic material body cannot be produced by hydrostatic tension or compression alone. Under the action of an isotropic force field the body decreases or increases in volume until the external forces are balanced by the (isotropic) internal reactions and a new position of equilibrium is established. On removal of the forces the body completely recovers its initial volume.

If the material body is made up of randomly arranged particles or of an arrangement of spherical symmetry, the equilibrium distances between the particles are changed by the external isotropic force field until the repulsive and attractive interatomic or intermolecular forces balance the external forces; such changes are transient, and no changes take place in the relative positions of the individual particle by which their energy would be reduced permanently. Although the observed practically perfect reversibility of volumetric deformation strongly supports the foregoing interpretation of the effect of an isotropic force field on an ideal statistically isotropic and homogeneous material, the existence, under such conditions, of inelastic deformation, no matter how small, cannot be wholly excluded for real materials. Several particles, no matter how few, will always be in positions from which stabilizing place changes are sufficiently probable actually to take place, since it is unlikely that under the action of an

external force field all particles of the body remain in conditions of perfect stability. Although the extent of local inelastic deformation in a statistically isotropic material body subject to an isotropic force field may be so small as to be imperceptible, such effects are bound to exist; being negligible under most conditions, they will manifest themselves under special conditions, for instance in the damping of volumetric vibrations.

In crystalline bodies the inherent anisotropy of the arrangement of particles produces different deformational responses in different directions. Hence, spherical symmetry of the external force field does not produce spherically symmetrical deformation, unless the lattice is one of spherical symmetry. The imposition of a state of hydrostatic pressure or tension on a crystalline body therefore, does not produce only volume changes but also changes of shape, which will be the more marked, the more pronounced the anisotropy of the atomic structure of the body. In crystals inelastic effects associated with isotropic force fields may therefore not be negligible. Such effects are practically nonexistent in the three-dimensionally symmetrical cubic crystals; they are, however, very pronounced in hexagonal crystals. Comparative compressibility measurements of crystals of various metals support this conclusion. Thus the difference in two directions of the compressibility of hexagonal crystals is very large; for zinc, a ratio of 1 to 7 between minimum and maximum compressibility has been observed.<sup>21-1</sup>

In addition to the anisotropy inherent in the crystal lattice there is another type of anisotropy, the existence of which may be the cause of inelastic deformation of both crystalline and amorphous bodies subject to an isotropic force field superimposed on a deviatoric (distortional) force field by which the interatomic distances in the different directions are changed by different amounts, proportional to the principal elastic strains. Different force fields necessarily produce different degrees of *strain anisotropy* in the initially isotropic structure. The strain anisotropy is clearly visible in the statistically isotropic photoelastic materials; the photoelastic response is but an expression of this phenomenon. Since the effect of strain anisotropy on deformation does not differ from the influence of inherent anisotropy, the superimposition of an isotropic force field on a deviatoric force field acting on a material body will produce not only volumetric

deformation but also a certain change of shape, the extent of which decreases with the intensity of the isotropic force field and the degree of strain anisotropy.

The generally assumed mutual independence of volumetric deformation and of change of shape of a material body under the action of external forces, therefore, can be considered but as a first approximation. As such it is of great value in simplifying the phenomenological analysis of deformation. Under certain conditions, however, such as superimposed very high hydrostatic pressure, this approximation may no longer be ade-

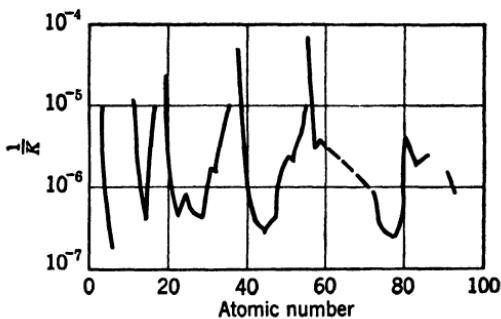


FIG. 21·1 Compressibility as function of atomic number.

quate to describe correctly the actual behavior of the material body.

Although volumetric deformation will not perceptibly deviate from elasticity, this elasticity is not linear, as may be inferred from Fig. 9·9. Whereas the decrease of the slope within the range of increasing distances between particles (extension) is usually of no particular significance, since fracture as a rule intervenes before the displacement reaches values at which the decrease in the slope is appreciable, its steep increase in the range of decreasing distances (compression), due to the sharply increasing repulsive energy of the particles, is responsible for the decrease, with increasing pressure, of the compressibility of all homogeneous materials; such decrease is expressed by an apparent increase, with decreasing volume, of the elastic modulus of volume change, usually designated as the *bulk modulus*.

The compressibility of elements varies periodically with their atomic number. This can be seen from Fig. 21·1 in which compressibilities have been plotted to a certain scale against atomic

number. A comparison with Figs. 8·2 and Fig. 8·3 shows the inverse and the parallel periodicities, respectively, of the compressibility versus atomic radius and of the density versus atomic radius relations. The most compressible solids are those that have the largest atomic radii and the lowest density; these are the alkali metals. Compressibility of a material may thus be visualized in terms of the compression of an aggregate of impenetrable "elastic" spherical electron shells. The larger the radii of the individual spheres, the easier and the more extensively can this body be compressed by hydrostatic pressure.

With increasing density and increasing external pressure the resistance to inelastic deformation increases in both crystalline and amorphous materials. In crystals this increase is expressed by an increase of the critical shear stress; this increase appears to be negligible at moderate pressures, but becomes appreciable at high pressures.<sup>21·2</sup> In amorphous materials the velocity of inelastic deformation is considerably reduced by superimposed pressure due to the increase of the coefficient of viscosity with increasing hydrostatic pressure.<sup>21·3</sup>

## 22. Fracture

Separation in a material body may be the result of shearing in two along a slip plane or of separation by extension normal to the separation plane. The forces and stresses that produce the first type of separation are the same that produce slip; thus the shear separation process is but a continuation of the slip process, extending beyond the limits of the crystal. The extensions producing true separation along planes normal to the lines of action of the forces, however, are unrelated to slip phenomena. The two separation processes are essentially of a different nature.

Genuine shearing separation resulting from the finite size of the body can take place, however, only in metals, in which the interatomic bonds are continually re-established in the course of the slip process. In ionic and homopolar crystals slip cannot proceed without a partial disruption of bonds, unless the thermal activation of the atoms is particularly strong and makes the reestablishment of the bonds possible. Such conditions are the exception, rather than the rule. In amorphous materials in which slip does not occur, separation is always associated with a force or an extension normal to the separation plane. That

such separation may be preceded and accompanied by substantial inelastic deformation does not change the fact that separation and deformation are different although not independent processes.

With the sole exception of the pure shearing into two of a single metal crystal the initiation of fracture represents a discontinuity in the deformation process. The change of state, which continuously proceeds within the material body subjected to forces and which manifests itself by the deformation, reaches a limit at the point of fracture, which is a condition of instability of the internal structure. When this condition is reached, the continuity of the transformation from mechanical (potential) into heat energy is broken by the appearance of a new mechanism of transformation of the energy of the applied forces into the energy of the newly created surfaces along which separation proceeds. The manner in which this transition point is reached depends essentially on the process of change of state, that is, of the deformation preceding and leading up to it.

Fracture is a process of progressive separation of atomic bonds which starts at points where the alternative mechanism of place change of particles is not available for the release of the potential energy of the applied forces, accumulated in the interatomic bonds, and where the absolute value of the accumulated bond energy is highest. The disruption of a number of atomic bonds increases the energy accumulated in the remaining bonds, since the energy of the acting forces is balanced by a further displacement of the bound particles from their position of equilibrium towards energy levels at which separation becomes increasingly possible. Atomic fracture is thus a "chain reaction" process; it may be very rapid and lead to macroscopic fracture under a single force application of short duration, if the energy of the applied forces is relatively high or if its distribution over the interatomic bonds is highly nonuniform; or macroscopic fracture may be slowly progressive and proceed under either repeated application or long duration of forces of a relatively low energy potential if the process is accompanied by a gradual change of the distribution of bond energies and by concentration of the response to the external forces within a decreasing number of bonds.

Since the initiation and propagation of fracture within the material depends on the level of potential energy stored up within

a number of excessively strained atomic bonds, it appears reasonable to assume that the occurrence of fracture on a macroscopic scale will be related to the accumulation of a certain critical amount of potential energy  $\Phi_s$ , which can not be dissipated into heat energy before it is released by separation.

The level of energy of the applied forces at which the critical state is reached depends necessarily on the process of energy dissipation preceding it, that is, on the relative amounts of stored-up and dissipated energy: the energy that is not stored up reversibly produces structural changes in the material, which are associated with its irrecoverable deformation. The close interrelation between fracture and inelastic deformation is the result of the fact that the greater the portion of the energy of the applied forces that is dissipated into heat energy in the course of inelastic deformation, the smaller the stored-up energy potential that remains available for fracture. Thus, under conditions that are not conducive to the dissipation of applied energy into heat energy, the main portion of the energy of the applied forces will be used to produce fracture. Such conditions are created by low temperatures, high rates of application of forces or of energy, as well as by a large share of the work of volumetric expansion in the total amount of the work of the applied forces. The difference in the effect on fracture of volumetric expansion and compression is due to the fact that, although volumetric energy cannot be effectively dissipated, volumetric compression evidently cannot cause separation. Elevated temperatures, low rates of application of force or of energy, and a large share of distortional work in the total work of the applied forces lead to a considerable intensification of the process of dissipation of applied energy into heat by place changes of particles, and thus increase the permanent deformation which precedes and accompanies fracture.

Fracture may start when, under the external conditions represented by forces and temperatures, not all the particles, of which the material body is built up, can find locations in which they are in a state of stable equilibrium. However, if such an unstable particle, in incessant thermal motion in a small region about its momentary position of equilibrium and continually exploring the potential field in which it is located, is able, in response to the imposed external conditions, and in order to restore its stability

reduced by these conditions, to find an alternative and more stable position of lower potential energy within the potential field of the forces of the surrounding particles, no atomic bond is disrupted. Evidently such change of place is possible only if (a) an alternative position within the potential field of the forces of the surrounding group of particles exists, and (b) the thermal activation of the particle is high enough to make this place change sufficiently probable to occur during the period of load application. Place changes of individual particles in response to external forces are therefore less probable at low than at high temperatures.

If the instability of the position, in which the particle finds itself as a result of changes in the external condition, cannot be resolved by a place change, the particle will tend to regain its stability by disrupting the bonds connecting it with certain of the surrounding particles, initiating fracture on the atomic scale. Whether and how rapidly this atomic fracture (which is a local instability phenomenon) will multiply and spread into a visible crack depends on energy considerations, since local instability spreads only if by this process the potential of all the forces involved is diminished.

Although the probability of occurrence of local bond disruption within the submicroscopic structure is highest at some intrinsic point of instability such as an imperfection of this structure, for instance a single dislocation or a concentration of dislocations in a block boundary, there are factors of instability present in any structural arrangement, even in the absence of imperfections. These factors are related to the temperature oscillations within the atomic or molecular structure, as a result of which transient cracks of atomic size might be opened during particularly large low-temperature energy fluctuations. It is, however, equally probable that compensating fluctuations of thermal energy above the mean would occur rapidly enough to *heal* these incipient *atomic cracks*. In fact, the heat energy set free at the same time at which atomic fracture occurs provides the heat just in the location where it is most needed to heal those fractures. Under such conditions the levels of potential energy pertaining, respectively, to the *cracked* and *uncracked* state differ by such infinitely small amounts that no definite trend in the separation process can develop. A clear trend towards large-scale propagation of

atomic fracture can develop only if the potential energies of the two states differ by finite amounts. Thus, if the interatomic distances are increased by the action of external forces, the gain in potential energy accompanying a spreading crack increases, as a result of the increased energy difference between the state of the material body before and after fracture.

Fracture cannot spread at the expense of the energy of the atomic forces, if the material is compressed so much that the interatomic forces are, on the average, repulsive. Under such conditions the opening of a crack would require the compressing of the particles on both sides of the potential crack, accompanied by an increase of potential energy. However, it is not impossible to set up such conditions that the increase in the energy of the internal forces associated with the spreading of a crack *against* compression is accompanied by such a decrease of the energy of the external forces that fracture may still occur, as the total potential of the system decreases. Such conditions which require a particular disposition of the external forces have been devised by P. W. Bridgman in some of his experiments on materials under the action of high hydrostatic pressure<sup>22·1</sup>.

Fracture cannot be produced by hydrostatic compression alone, as all particles are crowded together, not separated. If it is assumed that bond disruption is associated with a critical separation distance between the particles, a state of hydrostatic compression will increase the amount of the relative separation of particles necessary to cause bond disruption. A superimposed hydrostatic pressure, therefore, will increase the intensity of the forces required to produce fracture, whereas a hydrostatic tension will have the opposite effect. For moderate pressures, which do not cause changes in the interatomic distances of a magnitude comparable to the separation distance, these effects may be negligible. They will, however, become the more pronounced, the higher the volumetric compression as well as the compressibility of the material.

Materials in which even a moderate change of hydrostatic pressure or tension appreciably affects fracture under general conditions of stress are the microscopically and macroscopically inhomogeneous materials. Microscopic inhomogeneities of the structure produce a nonhomogeneous response within the material to the (homogeneous) state of hydrostatic stress. This

fact explains the considerable influence on the condition of fracture of very moderate spherical stresses in such materials as cast iron, concrete, and stone, in comparison to their negligible effect in the statistically homogeneous and relatively incompressible metals and amorphous substances.

Fracture being essentially the propagation from the submicroscopic into the macroscopic scale of instabilities within the potential field of the atomic or molecular forces, all effects that facilitate or inhibit the formation and propagation of such instabilities are of primary importance in influencing, if not determining, the occurrence and progress of fracture on a macroscopic scale.

The most important of the effects favoring the initiation of bond disruption are therefore the intrinsic inhomogeneities in the potential energy field of the interatomic or intermolecular forces as well as those produced by deformation. They cause the variation, over a wide range, of the energy levels of the particles and thus of the *energy contents* of the bonds between the particles. In statistically homogeneous and isotropic materials this variation of energy levels results from the statistical distribution of particles within the space occupied by them. In single crystals variation of energy levels of individual particles is theoretically nonexistent because of the uniform repetitive pattern of the distribution of particles in the ideal lattice. A relatively small slip however is sufficient to disturb the near homogeneity of the real lattice structure and to produce, in the glide planes, distorted atomic layers containing bonds of high and highly variable energy content. In two-phase or polyphase materials the inhomogeneities within the different phases will be of different order and type, since they are the result of the textural stresses which are of different orders of intensity. It is therefore primarily the inhomogeneity of the energy levels of particles within or near the boundaries of the individual constituent phases and within atomic layers distorted by deformation that is responsible for the actual separation strength of real two-phase or polyphase materials.

The most important effect counteracting the initiation and propagation of bond disruption is the thermal activation of particles. The more intense this activation, the higher the probability that a particle in unstable position will reach a more stable position by the mechanism of place change instead of by the

mechanism of bond disruption. The interplay between these two mechanisms is responsible for the character of the observable fracture. The influence of temperature on the fracture stress is relatively small, since this stress depends essentially on the interatomic force that defines the critical separation distance. Temperature, however, effects the relation between the *probabilities* of separation and of place change by which the *occurrence* of fracture and the extent of the preceding inelastic deformation is determined.

Attempts to compute the fracture stress from the interatomic forces<sup>22-2</sup> have proved as futile as the attempts to compute the critical shear stress in this manner, and for the same reason. Slip and fracture in real materials are determined by the *inhomogeneities* of the internal structure, not by the *regularities* of an assumed ideal arrangement. In crystals slip and fracture originate at dislocations, block boundaries, or other locations of structural inhomogeneity and proceed by the formation of further inhomogeneities and imperfections. Hence, both the critical shear stress and the fracture stress are determined by the relatively small number of *anomalously located* particles, rather than by the large majority of regularly arranged particles. It is, therefore, not surprising that the separation strength computed on the assumption of an ideal arrangement is between two and three orders of magnitude higher than the observed cohesive strength.

However, there is, at least in crystals, a definite effect of the mean energy of the atomic bonds on the trend of the observed separation strength. As this value depends on the initial imperfections within the crystal lattice, it will be the higher, the smaller the extent of such imperfections. It has been pointed out previously (see Art. 13) that the perfection of a lattice increases with the intensity of the interatomic forces: large forces may be sufficiently strong to overcome effectively the rapid reduction in the mobility of the particles in the solidification process due to the decreasing temperature, forcing the particles to form a regular lattice at a time when the effect of small forces would already have been blocked. Thus, the higher the intensity of the interatomic forces, that is, the higher the average lattice energy of the material, the less the imperfections in the atomic structure and the higher the real cohesive strength. This conclusion is borne out by all observations.

In crystals separation without preceding slip (*cleavage*) takes place along well-defined crystallographic planes called *cleavage planes*; these are frequently but not always identical with the slip planes. The stresses required to produce fracture by cleavage are different for different planes. Cleavage fractures result in the formation of flat plane surfaces; in metals they can be produced only if the force is applied almost perpendicularly to the potential cleavage plane, as otherwise slip interferes. Cleavage can be produced more easily at low temperatures, although the cleavage stress itself appears to be practically independent of temperature. However, the reduction of the velocity of slip with temperature, and the consequent increase of the critical shear stress increases the component of the force in the slip plane which would be necessary to initiate slip before cleavage is likely to occur. The deviation of the direction of the acting force from the direction perpendicular to the cleavage plane is, therefore, the less effective in preventing cleavage, the lower the temperature.

Fracture in polyphase molecular materials, particularly in high polymers, is initiated either by disruption of the primary bonds forming the individual chain molecules or by the pulling apart of intertwined groups of molecular chains, connected by relatively weak intermolecular forces, without any chain segment being broken. Materials in which fracture is governed by the latter mechanism, in general, will show considerably lower cohesion than materials breaking by rupture of individual chain segments, as the cohesive forces between segments are of atomic order of magnitude. However, if the chains are very densely spaced, the number of weak intermolecular bonds may increase so considerably that, in spite of the weakness of the individual bond, the total force required to produce fracture by the pulling apart of the intertwined chains becomes higher than the force required to break the primary bonds of chain segments. Since density of spacing of chain molecules is related to molecular weight, an increase of strength with the molecular weight of the material should be expected; the existence of such a relation has been confirmed by experiment.<sup>22-3</sup>

### References

15·1 R. C. TOLMAN, *Principles of Statistical Mechanics*, Oxford Univ. Press (1938) 71, 79.

15·2 S. GLASSTONE, K. J. LAIDLER, and H. EYRING, *The Theory of Rate Processes*, McGraw-Hill Book Co., New York (1941).

15·3 M. BORN, *Atomic Physics*, Blackie & Son, London (1945) 206.

15·4 P. DEBYE, *Ann. Physik* **39** (1912) 789.

16·1 J. A. HEDVALL, *Reaktionsfaehigkeit fester Stoffe*, J. A. Barth, Leipzig (1938) 106.

16·2 *Conf. Strength of Solids, Bristol* 1947 *Phys. Soc. London* (1948) 30 (Cottrell), 38 (Nabarro).

16·3 J. MUIR, *Phil. Trans.* **193** (1900) 1.

17·1 T. S. KE, *J. Applied Phys.* **20** (1949) 274.

17·2 W. ROSENHAIN, *J. Iron & Steel Inst.* **70** (2) (1906) 189.

18·1 H. J. GOUGH, *Proc. ASTM* **33/II** (1933) 3.

18·2 R. D. HEIDENREICH and W. SHOCKLEY, *Bristol Conf.*, op. cit., 57.

18·3 R. G. TRENTING and R. M. BRICK, *Trans AIME* **147** (1942) 128.  
N. DA C. ANDRADE and R. ROSCOE, *Proc. Phys. Soc.* **49** (1937) 152.

18·4 G. I. TAYLOR, *J. Inst. Metals* **62** (1938) 307.

18·5 A. KOCHENDOERFER, *Plastische Eigenschaften von Kristallen u. metallischen Werkstoffen*, J. Springer, Berlin (1941) 10.

18·6 F. SEITZ, *The Physics of Metals*, McGraw-Hill Book Co., New York (1943) 104.

18·7 J. H. HOLLOWOM and L. D. JAFFE, *Ferrous Metallurgical Design*, John Wiley & Sons, New York (1947) 40.

19·1 C. W. MACGREGOR and J. C. FISHER, *J. Applied Mechanics* **13** (1946) A-11.

19·2 *1st Rept. on Viscosity & Plasticity*, *Roy. Acad. Sci. Amsterdam* (1935) 208.

19·3 L. PRANDTL, *Z. angew. Math. & Mech.* **8** (1928) 85.

19·4 A. NADAI, *v. Kármán Anniversary Volume*, California Inst. Technology (1941) 237.

19·5 J. W. FREDERICKSON and H. EYRING, *Metals Technol.* (1948) T.P. 2423.

20·1 T. ALFREY, *Mechanical Behavior of High Polymers*, Interscience Pub., New York (1948) 340.

20·2 C. F. ELAM, *The Distortion of Metal Crystals*, Oxford Univ. Press (1935) 53.

20·3 R. KARNOP and G. SACHS, *Z. Physik* **41** (1927) 116.

20·4 O. HAASE and E. SCHMID, *Z. Physik* **33** (1925) 413.

20·5 W. BOAS, *Physics of Metals and Alloys*, John Wiley & Sons, New York (1947) 89.

20·6 A. W. McREYNOLDS, *J. Metals (Metals Trans.)* **1** (1949) 32.

20·7 D. HANSON and M. A. WHEELER, *J. Inst. Metals* **45** (1931) 229.

21·1 E. SCHMID and W. BOAS, *Kristallplastizitaet*, J. Springer, Berlin (1935) 199.

21·2 P. W. BRIDGMAN, *Rev. Modern Phys.* **17** (1945) 8.

21·3 P. W. BRIDGMAN, *Proc. Am. Acad. Arts Sci.* **61** (1925) 86.

22·1 P. W. BRIDGMAN, *J. Applied Phys.* **9** (1938) 517.

22·2 A. SMEKAL, *Physik. Z.* **34** (1933) 633.

22·3 T. ALFREY, op. cit., 493.

P A R T

B

## Mechanics of Inelastic Deformation



# CHAPTER

## 4

### MECHANICAL VARIABLES

#### 23. Mechanical State

The mechanical properties of a material are generally expressed in terms of the mechanical reactions to loads and to other external influences, such as temperature. These reactions are the observable and measurable expressions of the processes of exchange of mechanical energy and of its transformation (dissipation) into thermal energy. They can, therefore, be described by introducing the thermodynamical characteristics of the material body as functions of the *mechanical variables* (forces, displacements, and the like) and by applying the fundamental laws of thermodynamics. Mechanical properties in the widest sense are thus defined by four-dimensional relations between the mechanical and the thermodynamical variables, that is, between forces and displacements or between their specific values (stresses and strains), and between time and temperature. In the usually assumed isothermal conditions the relations are three-dimensional, that is, between forces, displacements (or stresses and strains), and time, temperature being simply a parameter. Relations connecting the mechanical and thermodynamical variables are called (mechanical) *equations of state*.

The existence of an equation of state implies that the force necessary to produce a certain deformation of a material body during a certain time and at a given temperature depends only on the momentary values of the deformation and its time derivatives and the momentary temperature, not on the extent and rate of previous deformation or temperature. The immedi-

ate effect of temperature in producing an intensification of the heat-energy fluctuations of the particles of the material body does not in itself interfere with the existence of an equation of state; although it changes the *shape* of the force-deformation-time relation, it does not affect the *continuity* of the relation. It is only under certain conditions that temperature effects also produce significant structural changes within the material (thermal recovery, grain growth, recrystallization, and precipitation). Since the rate and extent of those changes depend on the magnitude of previous deformation and on the conditions in which it was produced, the structural temperature effects seriously interfere with the existence of a mechanical equation of state by making the relations between the variables dependent on previous history.

Considering that real materials are generally not in a state of perfectly stable thermal equilibrium and that every irrecoverable deformation is the result of structural (atomic) changes within the material, *an equation of state cannot, in principle, be expected to exist*. However, for conditions for which the occurring structural changes do not appreciably affect the mechanical behavior, the existence of an equation of state can be assumed as a first approximation.

The assumption of the existence, for a certain material, of a mechanical equation of state makes it possible, by defining the relations existing at various temperatures between the *kinematical* and the *dynamical* variables to predict the mechanical behavior of the material under general conditions on the basis of a comparatively small number of tests under special conditions. The constants appearing in the equation of state are the mechanical constants of the material.

At constant temperature a mechanical equation of state can be represented by a three-dimensional curved surface, the dimensions being force or stress, deformation, and deformation velocity. It can be established either on the basis of structural (atomic) considerations (see Art. 19) or by the phenomenological procedure of performing mechanical tests. However, this latter method does not in itself establish the existence of an equation of state, unless additional observations are made to ascertain whether at constant temperature and given values of deformation and

deformation velocity the observed forces or stresses are one-valued and do not change with time.

Mechanical laws describe the deformation of material bodies by establishing relations between the dynamics and the kinematics of the deformation in terms of the mechanical variables. These variables must therefore belong to either of two groups: (a) the groups of dynamical variables, and (b) the group of kinematical variables. Since, phenomenologically, the deformable body is considered as a homogeneous isotropic continuous medium, all mechanical changes of a material body of arbitrary shape can be derived from those of the volume element by integration over space; therefore, the phenomenological study of mechanical behavior can be limited to the analysis of the mechanical changes produced in the volume element.

The state of deformation of a volume element is defined by its space coordinates and by the time derivatives (velocities, accelerations) of those coordinates. These magnitudes form the group of the kinematical variables.

There are three basic types of motion: translation, rotation, and pure deformation, that is, deformation without rotation. Pure deformation is again subdivided into volumetric change and volume-constant distortion. Any general motion of the volume element can thus be produced by a combination of translations, rotations, and pure deformations.

The dynamical state of a volume element is described by the dynamical variables, which are defined by relating each type of motion to a dynamical magnitude. Hence, translations are related to the forces by which they are produced, rotations to moments and deformations to stresses. According to the subdivision of deformations, stresses are subdivided into *isotropic stresses* producing volume changes and *deviatoric stresses*, that is, stresses producing distortions. These dynamical magnitudes, defined by their relation to the three basic types of motion, together with their derivatives in space and time, form the group of the dynamical variables. In defining those variables for the volume element, the forces and moments should be considered as *densities* of forces and moments.

The mechanics of translations and rotations is fully described by Newton's laws of motion which, for the volume element  $dV$

can be written in the form:

$$\Sigma f = \rho a dV \quad (23 \cdot 1)$$

for the dynamical equilibrium of forces  $f$ , and

$$\Sigma m = \Sigma f r = a r \rho dV \quad (23 \cdot 2)$$

for the dynamical equilibrium of moments  $m$ . In these equations  $r$  denotes the distance of the force vector from the center of  $dV$ ,  $a$  the acceleration, and  $\rho$  the mass density. The analysis of the motion is made possible by the application of d'Alembert's or Hamilton's principles. The only mechanical constant, the mass density  $\rho$  is sufficient to describe the mechanical behavior of the volume element.

In the mechanics of deformation d'Alembert's principle alone, relating forces and moments with translations and rotations, is insufficient to describe the phenomena; the necessary relation between the third pair of variables, the stresses and deformations, is provided by the mechanical equation of state. The number of material constants appearing in this equation will necessarily depend on the complexity of the behavior of the considered material, that is, on the number of constants contained in the equation of state.

It is important to define material constants in such a manner that they are independent of the conditions of the experiments from which they have been determined and that they describe the behavior of a volume element. Experiments provide, in general, relations between quantities observed on bodies of finite volume and specific shape; the immediate validity of these relations and of the parameters appearing therein is limited to the conditions of the experiment. These parameters become true mechanical constants only after the relations have been transformed so as to be valid for the volume element.

## 24. Dynamical Variables

Forces acting on the volume element are of two different types: (a) forces proportional to the mass or density of the element, which are the forces of gravity and of inertia, and (b) forces acting on the surface of the volume element, which are usually called *tractions* or *traction forces*. The distribution of the forces within every portion of the material body is governed by the

principle of dynamical equilibrium between the forces acting on the surface of the considered portion and the forces of gravity and inertia, acting on its mass. According to this principle the concept of *stress*, or more accurately, *tensor<sup>24·1</sup> of stress* is introduced for every infinitely small volume element; it expresses the fact that a surface of second degree (quadric) can always be found to represent the distribution over the unit sphere of the forces that act across all possible planes through the center of the element. *Stress* specifies the traction forces  $t$  acting across every given plane; it thereby determines two mutually perpendicular directions corresponding to the normal stress component  $t_n$ , which acts perpendicular to the plane, and the tangential or shear stress component  $t_s$  which acts within the plane. In the stress quadric three *principal* planes, at right angles to one another, exist across which the traction forces are normal, so that the tangential components  $t_s = 0$ . Three *principal axes* of stress are so specified that their directions coincide with those of the principal planes of the stress quadric; their respective lengths express the intensity of the traction forces per unit area across those planes.

Numerically, the state of stress within an infinitely small volume element, the rectangular (cartesian) coordinates of which are  $x_1$ ,  $x_2$ , and  $x_3$ , is completely defined by the specification of nine quantities, which are the components of the *stress tensor* in the directions of the axes of the coordinates  $x_1$ ,  $x_2$ ,  $x_3$ . If three surfaces of the volume element are defined by  $x_1 = \text{const}$ ,  $x_2 = \text{const}$ , and  $x_3 = \text{const}$ , respectively, the vectors of stress acting across those surfaces have the components  $(s_{11}, s_{12}, s_{13})$ ,  $(s_{21}, s_{22}, s_{23})$ , and  $(s_{31}, s_{32}, s_{33})$ , respectively, where the first subscript denotes the direction normal to the considered surface and the second the direction of the component. These are the nine components of the stress tensor  $\mathbf{T}_{jk}$  which can be written

$$\mathbf{T}_{jk} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \quad (24 \cdot 1)$$

This tensor of second rank defines in a concise manner the state of stress existing at a certain point within the material body.

The condition of force equilibrium 23·1 in the direction of the

three coordinate axes can be written in the well-known form:<sup>24·2</sup>

$$\frac{\partial s_{11}}{\partial x_1} + \frac{\partial s_{21}}{\partial x_2} + \frac{\partial s_{31}}{\partial x_3} = -f_1$$

$$\frac{\partial s_{12}}{\partial x_1} + \frac{\partial s_{22}}{\partial x_2} + \frac{\partial s_{32}}{\partial x_3} = -f_2 \quad (24 \cdot 2)$$

$$\frac{\partial s_{13}}{\partial x_1} + \frac{\partial s_{23}}{\partial x_2} + \frac{\partial s_{33}}{\partial x_3} = -f_3$$

where  $f_1, f_2$ , and  $f_3$  denote the components of the forces of inertia.

The condition that the volume element be in a state of moment equilibrium defined by eq. 23·2 requires that the stress tensor be symmetrical,<sup>24·3</sup> or that the tensor components,

$$s_{jk} = s_{kj} \quad (24 \cdot 3)$$

Tensor symmetry is an invariant property, so that eq. 24·3 holds under any coordinate transformation. Since the nine stress components of the tensor 24·1 are thus interrelated by the three eqs. 24·3, the state of stress at any point is completely determined by the six components of the symmetrical stress tensor,  $s_{11}, s_{22}, s_{33}, s_{12} = s_{21}, s_{23} = s_{32}$ , and  $s_{31} = s_{13}$ , which satisfy the equilibrium conditions and the given surface conditions in the three directions  $x_1, x_2, x_3$ .

FIG. 24·1 Stress vector on plane.

The components in the coordinate system  $(x_1, x_2, x_3)$  of the stress vector  $t$  on a plane the direction of which is defined by the three directional cosines  $l_1, l_2, l_3$  of its normal (Fig. 24·1) is obtained from the equilibrium condition:

$$t_1 = s_{11}l_1 + s_{21}l_2 + s_{31}l_3$$

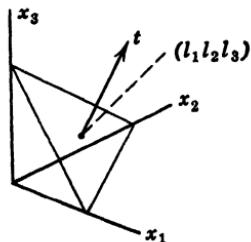
$$t_2 = s_{12}l_1 + s_{22}l_2 + s_{32}l_3 \quad (24 \cdot 4)$$

$$t_3 = s_{13}l_1 + s_{23}l_2 + s_{33}l_3$$

The vector  $t$  has a normal component  $t_n$  and a tangential component  $t_s$ , which, in terms of  $t_1, t_2, t_3$ , can be expressed by

$$t_n = t_1l_1 + t_2l_2 + t_3l_3 \quad (24 \cdot 5)$$

$$t_s^2 = t_1^2 + t_2^2 + t_3^2 - t_n^2 \quad (24 \cdot 6)$$



By expressing the values  $t_1, t_2, t_3$  in eq. 24·5 by eq. 24·4 and introducing the directional cosines  $l_1 = x_1/r, l_2 = x_2/r, l_3 = x_3/r$ , in terms of a radius vector  $\mathbf{r}$  through the origin in the direction of the normal to the plane, defined by the relation  $r^2 = \text{const}/t_n$ , the equation of the stress quadric is obtained:

$$s_{11}x_1^2 + s_{22}x_2^2 + s_{33}x_3^2 + 2s_{12}x_1x_2 + 2s_{23}x_2x_3 + 2s_{31}x_3x_1 = \text{const} \quad (24\cdot7)$$

If the axes  $x_1, x_2, x_3$  are rotated so as to coincide with the axes of the quadric (eq. 24·7), the mixed stress terms vanish. Hence,

$$t_1 = s_1l_1; \quad t_2 = s_2l_2; \quad t_3 = s_3l_3 \quad (24\cdot8)$$

where  $s_1, s_2, s_3$  denote the *principal stresses*, acting on the surface elements perpendicular to the principal directions of the stress quadric. When eq. 24·8 is introduced into eq. 24·5, the normal component,

$$t_n = s_1l_1^2 + s_2l_2^2 + s_3l_3^2 \quad (24\cdot9)$$

Hence, eq. 24·6 becomes

$$t_s^2 = s_1^2l_1^2 + s_2^2l_2^2 + s_3^2l_3^2 - (s_1l_1^2 + s_2l_2^2 + s_3l_3^2)^2 \quad (24\cdot10)$$

The values of the principal stresses are obtained if in eqs. 24·4 the vector components are expressed by eqs. 24·8, where  $s_1, s_2, s_3$  are considered the three possible values of the stress  $s$ . Considering that

$$l_1^2 + l_2^2 + l_3^2 = 1 \quad (24\cdot11)$$

four equations are thus obtained for  $l_1, l_2, l_3$  and the principal stress  $s$ . The determinant equation for  $s$  is obtained from eq. 24·4 in the form,

$$\begin{vmatrix} s_{11} - s & s_{21} & s_{31} \\ s_{12} & s_{22} - s & s_{32} \\ s_{13} & s_{23} & s_{33} - s \end{vmatrix} = 0 \quad (24\cdot12)$$

or explicitly,

$$s^3 - I_{s1}s^2 + I_{s2}s - I_{s3} = 0 \quad (24\cdot13)$$

For a symmetrical tensor 24·1, the following expressions represent the constant factors of eq. 24·13:

$$I_{s1} = s_{11} + s_{22} + s_{33} = s_1 + s_2 + s_3 \quad (24\cdot14)$$

$$\begin{aligned} I_{s2} &= s_{11}s_{22} + s_{22}s_{33} + s_{33}s_{11} - s_{12}^2 - s_{23}^2 - s_{31}^2 \\ &= s_1s_2 + s_2s_3 + s_3s_1 \end{aligned} \quad (24\cdot15)$$

$$\begin{aligned} I_{s3} &= s_{11}s_{22}s_{33} + 2s_{12}s_{23}s_{31} - s_{11}s_{23}^2 - s_{22}s_{31}^2 - s_{33}s_{12}^2 \\ &= s_1s_2s_3 \end{aligned} \quad (24 \cdot 16)$$

Equations 24·14, 24·15, and 24·16 express the three independent *invariants* of the stress tensor, which are the elementary symmetric functions of the principal stresses ( $I_{s3}$  represents the determinant of the stress components); they are called invariants, since they are invariant under any rotation of the coordinate system. The three values of the principal stress  $s_1, s_2, s_3$  are the roots of eq. 24·13.

It is evident that the foregoing three invariants are not the only possible forms of invariants, since by combining them algebraically new invariants can be formed.

A stress tensor can be resolved into its *isotropic* or *spherical* component and its *distortional* or *deviatoric* component. Hence,

$$\mathbf{T}_{jk} = \mathbf{T}_v + \mathbf{T}_{o_{jk}} \quad (24 \cdot 17)$$

where  $\mathbf{T}_v$  and  $\mathbf{T}_{o_{jk}}$  denote, respectively, the *spherical stress tensor* and the *stress deviator*, given by the expressions:

$$\mathbf{T}_v = \begin{pmatrix} p & o & o \\ o & p & o \\ o & o & p \end{pmatrix} \quad \text{and} \quad \mathbf{T}_{o_{jk}} = \begin{pmatrix} s_{11} - p & s_{21} & s_{31} \\ s_{12} & s_{22} - p & s_{32} \\ s_{13} & s_{23} & s_{33} - p \end{pmatrix} \quad (24 \cdot 18)$$

in which  $p$  denotes the mean hydrostatic stress,

$$p = \frac{1}{3}(s_{11} + s_{22} + s_{33}) \quad (24 \cdot 19)$$

The three invariants  $I_{0s1}, I_{0s2}$  and  $I_{0s3}$  of the stress deviator  $\mathbf{T}_{o_{jk}}$  in terms of those of the stress tensor are obtained by performing the operations indicated by eqs. 24·14 to 24·16 on the deviator 24·18. Hence,

$$I_{0s1} = 0 \quad (24 \cdot 20)$$

$$\begin{aligned} I_{0s2} &= I_{s2} - \frac{1}{3}I_{s1}^2 \\ &= -\frac{1}{6}[(s_{11} - s_{22})^2 + (s_{22} - s_{33})^2 + (s_{33} - s_{11})^2] \\ &\quad - (s_{12}^2 + s_{23}^2 + s_{31}^2) \\ &= -\frac{1}{6}[(s_1 - s_2)^2 + (s_2 - s_3)^2 + (s_3 - s_1)^2] \end{aligned} \quad (24 \cdot 21)$$

$$I_{0s3} = I_{s3} - \frac{1}{3}I_{s1} \left[ I_{s2} - \frac{2}{9}I_{s1}^2 \right] \quad (24 \cdot 22)$$

The invariants of the tensor or of the deviator of stress can be used to represent the state of stress or the state of deviatoric stress in a three-dimensional system of invariant coordinates ( $I_1, I_2, I_3$ ).

For plane stress with  $s_{33} = s_{13} = s_{23} = 0$  and within a direction defined by the cosines of the normal  $l_{1n} = \cos \alpha$  and  $l_{2n} = \sin \alpha$ , the components,

$$t_n = t_1 \cos \alpha + t_2 \sin \alpha \quad \text{and} \quad t_s = t_1 \sin \alpha - t_2 \cos \alpha \quad (24 \cdot 23)$$

are obtained from eqs. 24·5 and 24·6. Introducing the expressions 24·4 into eq. 24·23,

$$\begin{aligned} t_n &= s_{11} \cos^2 \alpha + s_{22} \sin^2 \alpha + s_{12} \sin 2\alpha \\ &= \frac{s_{11} + s_{22}}{2} + \frac{s_{11} - s_{22}}{2} \cos 2\alpha + s_{12} \sin 2\alpha \end{aligned} \quad (24 \cdot 24)$$

and

$$-t_s = \frac{s_{11} - s_{22}}{2} \sin 2\alpha - s_{12} \cos 2\alpha \quad (24 \cdot 25)$$

For

$$\tan 2\alpha = \tan 2\phi = \frac{2s_{12}}{s_{11} - s_{22}} \quad (24 \cdot 26)$$

the component  $t_s$  vanishes and  $t_n$  reaches the extreme value of the principal stress,

$$s_{1,2} = \frac{1}{2}(s_{11} - s_{22}) \pm \sqrt{\frac{(s_{11} - s_{22})^2}{4} + s_{12}^2} \quad (24 \cdot 27)$$

For

$$\tan 2\alpha = \tan 2\theta = -\frac{s_{11} - s_{22}}{2s_{12}} \quad (24 \cdot 28)$$

the shear component  $t_s$  reaches its extreme value,

$$t_{s\max} = \pm \sqrt{\frac{(s_{11} - s_{22})^2}{4} + s_{12}^2} = \pm \frac{s_1 - s_2}{2} \quad (24 \cdot 29)$$

while

$$t_n = \frac{s_{11} + s_{22}}{2} = \frac{s_1 + s_2}{2} \quad (24 \cdot 30)$$

Because  $\tan 2\phi \tan 2\theta = -1$ , the directions of principal shear stress are inclined under  $\pm 45^\circ$  to the direction of principal stress.

The planes of the principal shear stresses in the three-dimensional states of stress are obtained by eliminating from eq. 24·10

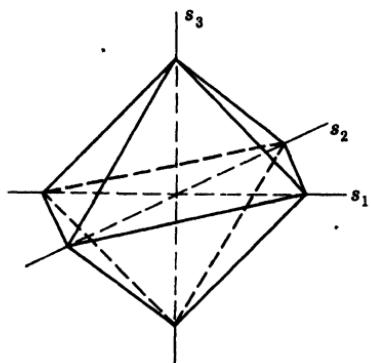


FIG. 24·2 Octahedron of directions of principal shear stresses.

the principal shear stresses form a regular octahedron whose corners lie on the axes of principal stress<sup>24·14</sup> (Fig. 24·2). Their values are

$$t_{s1} = \frac{1}{2}(s_2 - s_3); \quad t_{s2} = \frac{1}{2}(s_3 - s_1); \quad t_{s3} = \frac{1}{2}(s_1 - s_2) \quad (24 \cdot 31)$$

hence,

$$t_{s1} + t_{s2} + t_{s3} = 0 \quad (24 \cdot 32)$$

The normal and the shearing stresses  $t_{no}$  and  $t_{so}$  for any of the eight planes of the octahedron are called the *octahedral* normal and shearing stresses. Their values are

$$t_{no} = \frac{1}{3}[(s_1 + s_2 + s_3)] = \frac{1}{3}p$$

$$t_{so} = \frac{1}{3}[(s_1 - s_2)^2 + (s_2 - s_3)^2 + (s_3 - s_1)^2]^{1/2} \quad (24 \cdot 33)$$

This can be easily shown by computing  $t_n$  and  $t_s$  for any octahedral plane with the aid of eqs. 24·5, 24·6, and 24·8.

By substituting eq. 24·31 into the second eq. 24·33, the octahedral shear in terms of the three principal shear stresses:

$$t_{so} = \frac{2}{3}[t_{s1}^2 + t_{s2}^2 + t_{s3}^2]^{1/2} = \sqrt{-\frac{2}{3}I_{s02}} \quad (24 \cdot 34)$$

Because of eq. 24·32,  $t_s$  can be expressed in terms of two of the

one directional cosine with the aid of the relation 24·11 and subsequently differentiating eq. 24·10 with respect to the remaining cosines. By repeating this procedure with all three cosines, three pairs of extremal conditions are obtained for the planes of extreme values of shearing stress. They are fulfilled on the planes containing one axis and bisecting the angle between the other two axes of principal stress under  $45^\circ$ . The directions of

principal shear stresses only:

$$t_{so} = \frac{2}{3}[2(t_{s1}^2 + t_{s1} \cdot t_{s2} + t_{s2}^2)]^{1/2} \quad (24 \cdot 35)$$

## 25. Kinematical Variables

For the definition of deformation or displacement two positions of the material body, one initial and one terminal, must be considered. The displacement of the mass points of the material from one position to another shows a distribution in space that is necessarily governed by the principle of continuity, unless fracture occurs and this continuity is broken. According to this principle, the concept of *strain* or, more accurately, of a *strain tensor* expresses the fact that the relative displacements, that is, the changes in length in the immediate neighborhood of the considered point, are always so distributed over the unit sphere surrounding the point that a quadric, the *strain ellipsoid*, changes its form from a sphere in the initial unstrained position to an ellipsoid in the terminal position, while another quadric, the *reciprocal strain ellipsoid*, simultaneously changes from an ellipsoid into a sphere. In both the initial and the terminal positions, the strain ellipsoid and the reciprocal strain ellipsoid define, by their common center and their curves of intersection, a conical surface, termed the *cone of zero elongation*, which represents an invariant in the sense that its generating lines have the same length in both positions, as if they were rigid rods. Whereas the strain ellipsoid defines the state of strain under applied forces, the reciprocal strain ellipsoid defines the strain associated with the removal of forces.

Considering a series of parallel planes, these planes will glide over one another during the progress of the deformation; at the same time their distances apart in a direction normal to the planes will change, while in a direction at right angles to both directions the relative displacements are zero. Thus, for each plane two mutually perpendicular directions are defined as the directions of normal strain and of shear strain. There are three *principal planes* in the strain quadric at right angles to one another, and for these the shear strains vanish and the displacements are normal. The directions of the normals to the principal planes of the strain quadric are usually defined as the three main axes of strain or axes of *principal strain*.

The magnitude of these axes can be defined in various ways. Although, in principle, this definition is arbitrary, certain requirements must be met in order to obtain all relations in the simplest possible form. Generally, mechanical variables obey a law of simple superposition; values may be added algebraically to obtain the resulting value. In defining the magnitude of strain however, this possibility depends on the definition of the measure of strain.

There are two modes of definition of the deformation of a continuous medium. In terms of *Lagrangian coordinates* the independent variables are the positions  $x_i$  of the mass points in the initial unstrained state; the derivatives of the displacements are determined at these positions. In terms of *Eulerian coordinates* the independent variables are the displacements  $u_i$  which define the positions of the mass points in the deformed state, and the derivatives are determined at the positions  $(x_i + u_i)$ . When the displacements are so small that products of derivatives can be neglected, the difference between these two modes of defining strain vanishes.

When the deformations are not limited to infinitesimal values, different definitions of the measure of strain will necessarily lead to different stress-strain relations. Evidently a stress-strain relation that is linear in one strain measure will be nonlinear in every other measure. In principle, any arbitrary function of the elongation ratio of the axes of the strain ellipsoid  $\lambda_i$  can be used as a measure of strain, provided such a function is reduced to the classical measure,  $e_i = (\lambda_i - 1)$  for infinitesimal strain.

If a point of a material body with coordinates  $(x_1 x_2 x_3)$  is displaced by a distance  $u$ , the projections of the displacement vector in the directions of the coordinate axes are  $(u_1 u_2 u_3)$ ; since displacements must be continuous functions of the coordinates, the change in the components of the relative displacements of the neighboring points, under the assumption of small deformations can be written in the form,

$$du = \sum \frac{\partial u}{\partial x_i} dx_i \quad (25 \cdot 1)$$

where the derivations are taken at  $(x_1 x_2 x_3)$ . The change of displacement  $du$  of the considered point, that is, its deformation, can be specified by the nine components  $\partial u / \partial x_i$  of a nonsym-

metrical tensor of second rank; the tensor of the total deformation

$$\mathbf{U}_{jk} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{pmatrix} \quad (25 \cdot 2)$$

This is evidently not the tensor of strain, since it contains the rotation  $\omega_i = \frac{1}{2} \left( \frac{\partial u_j}{\partial x_k} - \frac{\partial u_k}{\partial x_j} \right)$ , whereas the tensor of strain, related to the tensor of stress of a volume element in dynamical equilibrium, must necessarily be symmetrical and therefore rotationless. Thus, a new symmetrical tensor must be developed, by splitting off the symmetrical part and taking the averages of the terms  $\partial u_j / \partial x_k$  lying symmetrically to the diagonal  $\partial u_i / \partial x_i$  of the tensor  $\mathbf{U}_{jk}$ . The components of this symmetric tensor are therefore defined by the expressions:

$$e_{ii} = \frac{\partial u_i}{\partial x_i} \quad \text{and} \quad g_{jk} = \frac{1}{2} \left( \frac{\partial u_j}{\partial x_k} + \frac{\partial u_k}{\partial x_j} \right) = g_{kj} \quad (25 \cdot 3)$$

where the terms  $e_{ii}$  denote the normal strains and  $g_{jk}$  the shear strains or angle changes. The three components of the vector of pure deformation are thus given by the equations:

$$\begin{aligned} du_1 &= e_{11} dx_1 + g_{12} dx_2 + g_{13} dx_3 \\ du_2 &= g_{21} dx_1 + e_{22} dx_2 + g_{23} dx_3 \\ du_3 &= g_{31} dx_1 + g_{32} dx_2 + e_{33} dx_3 \end{aligned} \quad (25 \cdot 4)$$

The symmetric strain tensor can therefore be written in the form,

$$\mathbf{E}_{jk} = \begin{pmatrix} e_{11} & g_{12} & g_{13} \\ g_{21} & e_{22} & g_{23} \\ g_{31} & g_{32} & e_{33} \end{pmatrix} \quad (25 \cdot 5)$$

where  $g_{jk} = g_{kj}$ . The deformation tensor of eq. 25.2 is thus made up of the symmetric tensor of strain (eq. 25.5) and of the asymmetric rotation represented by the vector components:

$$\begin{aligned}dr_1 &= -\omega_3 dx_2 + \omega_2 dx_3 \\dr_2 &= \omega_3 dx_1 - \omega_1 dx_3 \\dr_3 &= -\omega_2 dx_1 + \omega_1 dx_2\end{aligned}\quad (25 \cdot 6)$$

As in the case of the stress tensor, the three directions for which the mixed components of strain vanish are the axes of *principal strain*  $e$ ; the values of principal strain are obtained by solving eqs. 25.4 for  $du_i = e dx_i$ . The determinant equation, the three roots  $e_1, e_2$ , and  $e_3$  of which are the principal strains can therefore be written in the form,

$$\begin{vmatrix} e_{11} - e & g_{12} & g_{13} \\ g_{21} & e_{22} - e & g_{23} \\ g_{31} & g_{32} & e_{33} - e \end{vmatrix} = 0 \quad (25 \cdot 7)$$

or, explicitly,

$$e^3 - I_{e1}e^2 + I_{e2}e - I_{e3} = 0 \quad (25 \cdot 8)$$

where  $I_{e1}, I_{e2}, I_{e3}$  denote the three *invariants* of the strain tensor:

$$I_{e1} = e_{11} + e_{22} + e_{33} = e_1 + e_2 + e_3 \quad (25 \cdot 9)$$

$$\begin{aligned}I_{e2} &= e_{11}e_{22} + e_{22}e_{33} + e_{33}e_{11} - g_{12}^2 - g_{23}^2 - g_{31}^2 \\&= e_1e_2 + e_2e_3 + e_3e_1 \quad (25 \cdot 10)\end{aligned}$$

$$\begin{aligned}I_{e3} &= e_{11}e_{22}e_{33} + 2g_{12}g_{23}g_{31} - e_{11}g_{23}^2 - e_{22}g_{31}^2 - e_{33}g_{12}^2 \\&= e_1e_2e_3 \quad (25 \cdot 11)\end{aligned}$$

These expressions are invariant with respect to any orthogonal transformation, that is, rotation, of the coordinate system. The third invariant  $I_{e3}$  represents the determinant of the strain components.

If terms of higher than first order in strain are neglected, the first invariant of strain  $I_{e1}$  represents the dilatation of a unit volume  $V$ , since the specific volume change:

$$\begin{aligned}\frac{dV}{V} &= (1 + e_1)(1 + e_2)(1 + e_3) - 1 \\&= (e_1 + e_2 + e_3) + \text{terms of higher order in } e \quad (25 \cdot 12)\end{aligned}$$

Hence, in first approximation,

$$I_{e1} = e_1 + e_2 + e_3 \doteq \frac{dV}{V} = 3e_v \quad (25 \cdot 13)$$

Every strain tensor  $\mathbf{E}_{jk}$  can be resolved into its *volumetric component*, the *spherical tensor*  $\mathbf{E}_v$ , and its *distortional component* or *deviator*  $\mathbf{E}_{0jk}$ . Hence,

$$\mathbf{E}_{jk} = \mathbf{E}_v + \mathbf{E}_{0jk} \quad (25 \cdot 14)$$

where

$$\mathbf{E}_v = \begin{pmatrix} e_v & 0 & 0 \\ 0 & e_v & 0 \\ 0 & 0 & e_v \end{pmatrix} \text{ and}$$

$$\mathbf{E}_{0jk} = \begin{pmatrix} (e_{11} - e_v) & g_{12} & g_{13} \\ g_{21} & e_{22} - e_v & g_{23} \\ g_{31} & g_{32} & e_{33} - e_v \end{pmatrix} \quad (25 \cdot 15)$$

The three invariants of the deviator of strain  $\mathbf{E}_{0jk}$  in terms of those of the strain tensor are

$$I_{0e1} = 0 \quad (25 \cdot 16)$$

$$I_{0e2} = I_{e2} - \frac{1}{3}I_{e1}^2 = -\frac{1}{6}[(e_1 - e_2)^2 + (e_2 - e_3)^2 + (e_3 - e_1)^2] \quad (25 \cdot 17)$$

$$I_{0e3} = I_{e3} - \frac{1}{3}I_{e1}(I_{e2} - \frac{2}{3}I_{e1}^2) \quad (25 \cdot 18)$$

The state of strain can be represented in a plane of strain with the aid of Mohr's circles of strain in a similar way that the state of stress can be represented in Mohr's stress plane.<sup>25·1</sup> By replacing the six components of the stress tensor  $\mathbf{T}_{jk}$  by those of the tensor of strain  $\mathbf{E}_{jk}$ , similar relations are obtained between the unit sphere and Mohr's plane of strain to those obtained for Mohr's plane of stress. This analogy of the circles of stress and of strain is valid however only for small deformations, since the simple definition of shear strain (eq. 25·3) is not valid beyond this range.

For pure distortional strain defined by  $e_v = 0$  the equations for the octahedral shear strain may be expressed in analogy to the equations of octahedral shear stress:

$$g_o = \frac{2}{3}[(e_1 - e_2)^2 + (e_2 - e_3)^2 + (e_3 - e_1)^2]^{\frac{1}{2}} \quad (25 \cdot 19)$$

considering that the principal shear strains,

$$g_1 = e_2 - e_3; \quad g_2 = e_3 - e_1; \quad g_3 = e_1 - e_2 \quad (25 \cdot 20)$$

Thus eq. 25·19 may be written

$$g_o = \frac{2}{3}[g_1^2 + g_2^2 + g_3^2]^{1/2} = \sqrt{-\frac{8}{3}I_{0e2}} \quad (25 \cdot 21)$$

Because of  $e_v = 0$ , the sum  $(g_1 + g_2 + g_3) = 0$ ; hence,

$$g_o = \frac{2}{3}[2(g_1^2 + g_1g_2 + g_2^2)]^{1/2} = \frac{2}{3}[2(e_1^2 + e_1e_2 + e_2^2)]^{1/2} \quad (25 \cdot 22)$$

The kinematical state of a material body is not sufficiently defined by its initial and its deformed position but requires the consideration of the time during which the process of deformation takes place. This is done by introducing the velocity vector  $\mathbf{v}$  instead of the displacement vector  $\mathbf{u}$ . The symmetric tensor  $\dot{\mathbf{E}}_{jk}$  of the strain velocities is obtained by replacing in eqs. 25·3 and 25·5 the components of the strain vectors by those of the strain velocity vectors.

The classical definition of strain relates the elongation  $dl$  of the axes of the strain ellipsoid to their length  $l_0$  before deformation. This has the disadvantage that the strain obtained by an  $n$ -fold repetition of the same elongation  $dl$  is different from that resulting from the elongation  $n dl$  imposed in one operation. Because of this definition of strain with reference to the length before deformation, the same elongation,

$$dl = [l_1 - l_0] = [l_2 - l_1] = [l_3 - l_2] = [l_n - l_{(n-1)}] \quad (25 \cdot 23)$$

does not lead to the same value of strain. The strain associated with an  $n$ -fold repetition of  $dl$  is therefore

$$\begin{aligned} \sum_n e_i &= \sum_n dl/l_{(i-1)} = \left(\frac{l_1}{l_0} - 1\right) + \dots + \left(\frac{l_n}{l_{n-1}} - 1\right) \\ &= \frac{dl}{d_0} + \frac{dl}{l_1} + \dots + \frac{dl}{l_{n-1}} \neq \frac{n dl}{l_0} \end{aligned} \quad (25 \cdot 24)$$

This difficulty can be overcome by introducing a different definition of strain, which was first suggested by Roentgen and later by Ludwik<sup>25·2</sup> and more recently advanced by Hencky<sup>25·3</sup> on the basis of a systematic consideration of its mathematical implications. If strain is defined as the ratio of the momentary elongation  $dl$  to the momentary length  $l$  a finite change of length from  $l_0$  to  $l$  is associated with the strain:

$$\epsilon = \int_{l_0}^l \frac{dl}{l} = \ln \frac{l}{l_0} = \ln \left(1 + \frac{dl}{l_0}\right) = \ln (1 + e) = \ln \lambda \quad (25 \cdot 25)$$

This measure is called the *logarithmic* or *natural strain* and will be designated by  $\epsilon$ . Hencky has shown that the strains  $\epsilon$  form a *group* and that therefore they can be added algebraically. Hence,

$$\sum_n \epsilon_i = \sum_n \int_{l_{i-1}}^{l_i} \frac{dl}{l} = \ln(l_1/l_0) + \ln(l_2/l_1) + \dots \\ = \ln \frac{l_1 l_2 \dots l_n}{l_0 l_1 \dots l_{(n-1)}} = \ln(l_n/l_0) = \int_{l_0}^{l_n} \frac{dl}{l} \quad (25 \cdot 26)$$

The comparison of eq. 25·26 with eqs. 25·12 and 25·13 shows that, as a result of the logarithmic definition of strain, eq. 25·13 is no longer an approximation.

Since, according to eq. 25·13, the dilation  $3\epsilon_v$  is equal to the first invariant of the strain tensor if and only if the strain is logarithmically defined, unless the analysis is restricted to infinitesimal strains, it is only in the logarithmic measure of strain that the resolution of a strain tensor into a volumetric and a distortional component retains its physical significance beyond the range of infinitesimal strains.

In the general case of a triaxial pure (rotationless) deformation, logarithmic strain is defined by

$$\begin{aligned}\epsilon_1 &= \ln(1 + e_1) = \ln \lambda_1 \\ \epsilon_2 &= \ln(1 + e_2) = \ln \lambda_2 \\ \epsilon_3 &= \ln(1 + e_3) = \ln \lambda_3\end{aligned}\quad (25 \cdot 27)$$

where  $e_1, e_2, e_3$  denote the classically defined strains. Hence

$$\lambda_1 = e^{\epsilon_1}, \quad \lambda_2 = e^{\epsilon_2}, \quad \lambda_3 = e^{\epsilon_3} \quad (25 \cdot 28)$$

All relations established for  $e$  are equally valid for  $\epsilon$ . Thus, the principal shear strains  $\gamma_i$

$$\gamma_i = \epsilon_j - \epsilon_k = \ln(1 + e_j) - \ln(1 + e_k) \quad (25 \cdot 29)$$

The components  $\gamma_i$  fulfill the condition,

$$\Sigma \gamma_i = 0 \quad (25 \cdot 30)$$

and the logarithmic octahedral shear strain,

$$\gamma_o = \frac{1}{3}[2(\epsilon_1^2 + \epsilon_1\epsilon_2 + \epsilon_2^2)]^{1/2} \quad (25 \cdot 31)$$

The principle of algebraic addition of finite logarithmic strains

is valid, however, only for pure strains, that is, as long as the axes of the strain ellipsoid do not rotate during the deformation. These are the states of strain in which perpendicular systems of straight parallel lines are deformed in such a way that they remain perpendicular systems of straight and parallel lines. There are only three such states: (1) pure volume change without distortion, (2) unidirectional strain and (3) pure shear without volume change where a rhombus is changed into a congruent rhombus by interchanging the acute and obtuse angles as well

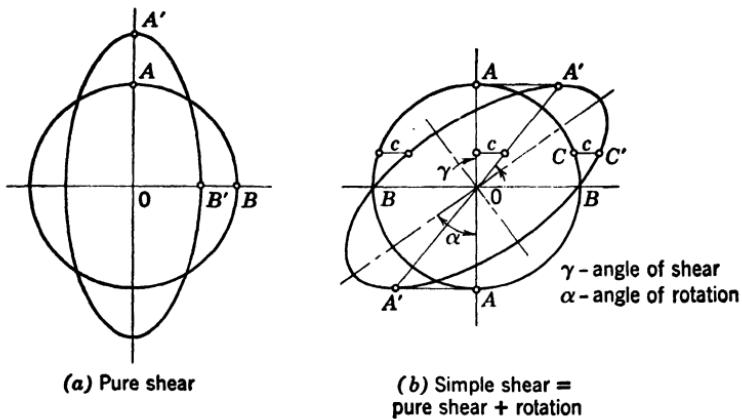


FIG. 25·1 Illustration of deformation of circle in (a) pure shear and (b) simple shear.

as all combinations of states 1, 2, and 3. It is important to remember that *simple shear* is not a pure strain but a combination of a pure strain and a rotation (Fig. 25·1).

The fact that the definition of strain used will necessarily affect the relations between the dynamical and kinematical variables observed in mechanical tests of materials is of considerable importance in the interpretation of test results. Examples of the differences in the plotted test results introduced by the arbitrariness of the definition of strain are given in Art. 87. It should be kept in mind that in attributing physical meaning to curves obtained in mechanical tests, the arbitrariness of the definition of strain requires careful consideration. Unless deformations are infinitely small, the definition of strain also affects the dynamical variables in finite bodies. Whenever the unstrained (initial) and strained shapes of the body cannot be

considered practically identical, as a result of the strain being finite, the action of the forces must be related to the strained shape in order to obtain the *true* stresses.

In the study of the relations between the kinematical and dynamical variables it is necessary to bear in mind the tensor character of these relations even if, as in this book, tensors are introduced with the purpose of familiarizing the engineer with the concept itself, and of having a short and therefore convenient expression for relations between general states of stress and strain, rather than of making use of the methods of tensor analysis. Physical characteristics, however, must be classified according to their *tensorial rank*, because in cases of a transformation of the reference system, the transformation of characteristics of different tensorial rank follows different laws. A physical constant, as any scalar magnitude, is a tensor of *zero rank*, represented by a number. A tensor of *first rank* is a vector which in the  $r$ -dimensional space is represented by  $r$  components; displacements, velocities, and forces are tensors of first rank. Tensors of second rank, such as stresses, strains, and strain velocities are represented by a matrix of  $r^2$  components.

It is a fundamental law of physics that no relation between variables is of general validity *unless the related characteristics are of equal tensorial rank*; if this condition is not fulfilled, a single transformation of the reference system is sufficient to break or alter the established relation, the physical meaning of which would therefore be dependent on and strictly limited to the specific conditions of the test from which it has been derived.

## 26. Mechanical Variables in Curvilinear Coordinates

In analyzing certain problems it is expedient to replace the rectangular system of coordinates by a curvilinear orthogonal system. The practically most important system is that of cylindrical coordinates which in the state of plane stress is identical with that of polar coordinates. The three directions are the radial direction  $r$ , the tangential direction  $\theta$ , and the vertical direction  $z$ .

The symmetrical stress tensor is given by the six stress components:

$$\mathbf{T}_{jk} = \begin{pmatrix} s_{rr} & s_{\theta r} & s_{zr} \\ s_{\theta \theta} & s_{zz} & \\ & s_{zz} & \end{pmatrix} \quad (26 \cdot 1)$$

The principal stresses  $s_r$ ,  $s_\theta$ ,  $s_z$  (axes of the stress ellipsoid) are obtained in the manner outlined in Art. 24 by solving the determinant equation; the invariants of the tensor are again the three elementary symmetric functions of the principal stresses. The equilibrium conditions are<sup>26·1</sup>

$$\begin{aligned}\frac{\partial s_{rr}}{\partial r} + \frac{1}{r} \frac{\partial s_{r\theta}}{\partial \theta} + \frac{\partial s_{rz}}{\partial z} + \frac{s_{rr} - s_{\theta\theta}}{r} &= -f_r \\ \frac{\partial s_{r\theta}}{\partial r} + \frac{1}{r} \frac{\partial s_{\theta\theta}}{\partial \theta} + \frac{\partial s_{\theta z}}{\partial z} + \frac{2s_{r\theta}}{r} &= -f_\theta \quad (26 \cdot 2) \\ \frac{\partial s_{rz}}{\partial r} + \frac{1}{r} \frac{\partial s_{\theta z}}{\partial \theta} + \frac{\partial s_{zz}}{\partial z} + \frac{s_{rz}}{r} &= -f_z\end{aligned}$$

where the  $f_i$  denote the components of the external forces.

If the position of a point in space is given by its three coordinates  $r$ ,  $\theta$ ,  $z$ , the line element  $ds$  is defined by

$$ds^2 = dr^2 + r^2 \cdot d\theta^2 + dz^2 \quad (26 \cdot 3)$$

The change of coordinates resulting from the deformation in terms of the components of the displacement vector  $u_r$ ,  $u_\theta$ , and  $u_z$  are:

$$dr = u_r; \quad d\theta = \frac{u_\theta}{r}; \quad dz = u_z \quad (26 \cdot 4)$$

The tensor of strain is defined by the change in the coefficients of eq. 26·3, and its components are obtained by evaluating the difference in length of the line element  $ds$  before and after deformation and by computing the components in the directions of the coordinates. The six components of the symmetrical strain tensor are given by the expressions:<sup>26·1</sup>

$$\begin{aligned}e_r &= \frac{\partial u_r}{\partial r}; & g_{r\theta} &= \frac{1}{r} \cdot \frac{\partial u_\theta}{\partial r} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r} \\ e_\theta &= \frac{1}{r} \cdot \frac{\partial u_\theta}{\partial \theta} + \frac{u_r}{r}; & g_{\theta z} &= \frac{\partial u_\theta}{\partial z} + \frac{1}{r} \cdot \frac{\partial u_z}{\partial \theta} \quad (26 \cdot 5) \\ e_z &= \frac{\partial u_z}{\partial z}; & g_{sr} &= \frac{\partial u_z}{\partial r} + \frac{\partial u_r}{\partial z}\end{aligned}$$

The specific volume change,

$$3e_v = e_r + e_\theta + e_z = \frac{1}{r} \cdot \frac{\partial}{\partial r} (ru_r) + \frac{1}{r} \cdot \frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} \quad (26 \cdot 6)$$

The principal strains (axes of the strain ellipsoid) and the invariants of the tensor and deviator of strain are obtained from eq. 25·8 and the eqs. 25·9 to 25·11 and 25·16 to 25·18.

In the practically important cases considerable simplifications of the general equations are introduced by the particular conditions of loading and of deformation.

### References

- 24·1 L. BRILLOUIN, *Les Tenseurs en mécanique et en élasticité*, Dover Pub., New York (1946) 212.
- 24·2 L. S. SOKOLNIKOFF, *Mathematical Theory of Elasticity*, McGraw-Hill Book Co., New York (1946) 41.
- 24·3 *Ibid.*, 42.
- 24·4 S. TIMOSHENKO, *Theory of Elasticity*, McGraw-Hill Book Co., New York (1934) 87.
- 25·1 H. W. SWIFT, *Engineering* **162** (1946) 381.
- 25·2 P. LUDWIK, *Elemente d. Technologischen Mechanik*, J. Springer, Berlin (1909) 17.
- 25·3 H. HENCKY, *Z. Physik* **55** (1929) 145; *Ann. Physik* **5(2)** (1929) 617.
- 26·1 I. S. SOKOLNIKOFF, *op. cit.*, 202.

C H A P T E R  
5

## THE MECHANICAL EQUATION OF STATE

### 27. Change of State. Thermodynamic Considerations

The coordinates in terms of which the mechanical state of a deformable material body is described are the kinematical variables and the dynamical variables. Changes of mechanical state are defined by continuous functions of these coordinates, which represent *equations of state*.

Changes of state may be either *reversible* or *irreversible*. Irreversible changes of state are accompanied by permanent (identifiable or unidentifiable) changes in the pattern of the internal structure of the material on the various levels of aggregation of structural elements. Changes of pattern accompanying reversible changes of state are transient, since the definition of reversibility implies the identity of the initial and final state. On a structural scale this is equivalent either to the identity or to the equal probability of occurrence of the initial and the final structural pattern. The continuity of the mechanical equation of state for a continuously changing independent variable, expressing continuity of change of the dependent variable, depends on changes of structural pattern being the exclusive result of changes of the independent mechanical variable. If the change of structural pattern is "spontaneous," that is, unrelated to changes of the independent variable, uncontrollable discontinuities in the equation of state must be the result.

Changes that occur spontaneously are manifestations of the fact that the considered system is not in equilibrium. Inversely, a criterion of equilibrium can be established with regard to the

fact that the properties of a system do not change spontaneously. Thus, an equation of state may be assumed to exist only if the momentary values of the dependent variable are fully determined by the momentary value of the independent variable and do not depend on the previous history, that is, on the path along which these values have been attained. In this case the changes of structural pattern accompanying the forced change of state must be either reversible or unidentifiable.

Essentially, equations of state describe processes of energy transformations expressed either in terms of forces or stresses and displacement or deformations or directly in terms of the different types of mechanical and heat energy. Such equations should therefore be derived from the fundamental principles governing processes of energy transformation, that is, from the two principal laws of thermodynamics.

The first law or law of conservation of energy for a mechanically closed (conservative) system states that the energy of this system is invariable. The change of energy of the system per unit time is therefore equivalent to the energy per unit time supplied to the system by the work of the external forces. Thus the work  $A_{12}$  necessary to transfer a system from the energy state  $W_1$  to the energy state  $W_2$  depends only on these two states and does not depend on the path over which this transfer has taken place. Hence,

$$W_2 - W_1 = A_{12} \quad \text{or, per unit of time,} \quad \frac{dW}{dt} = \frac{dA}{dt} \quad (27 \cdot 1)$$

In the form of eq. 27·1 the law of conservation of energy is limited to purely mechanical phenomena; the energy of the system  $W$  is the sum of the kinetic energy  $W_k$  and the internal or potential energy  $\Phi$ ; the work  $A$  is the work supplied by the external forces. If the mechanical law of energy conservation is extended to include the principle of equivalence of work and heat, the first law of thermodynamics for constant mass takes the form,

$$W_2 - W_1 = Q + A \quad \text{or, per unit of time,}$$

$$\frac{dW}{dt} = \frac{dW_k}{dt} + \frac{d\Phi}{dt} = \frac{dQ}{dt} + \frac{dA}{dt} \quad (27 \cdot 2)$$

where  $Q$  denotes the heat energy in mechanical units, applied to the system from outside. Equilibrium of a conservative system is defined by the absence of changes in kinetic, external, and thermal energy; hence,  $d\Phi/dt = 0$  and  $\Phi \rightarrow$  extremal value.

Since, according to the energy law of mechanics,

$$dW_k = dA - dW_i \quad (27 \cdot 3)$$

where  $W_i$  denotes the work of the internal forces (stresses), the relation,

$$\frac{d\phi}{dt} = \frac{dQ}{dt} + \frac{dW_i}{dt} \quad (27 \cdot 4)$$

is obtained from eqs. 27 · 2.

The equivalence of mechanical work and heat contained in the first law is subject to the limitation of the second law, which expresses the fact that the exchange of mechanical energy into heat is not a reversible process, at least not on a phenomenological scale. Only as far as quantity is concerned can work and heat be equated exactly, since they are both forms of energy subject to the first law; however the exchange from one form into another is noncommutative. In real mechanical processes the creation of heat by friction or friction-like processes can never be prevented; such processes can therefore not be perfectly reversible.

According to the second law of thermodynamics the irreversibility of a mechanical process is defined by the condition that this process or change of state be accompanied by an increase of the *entropy* of the considered mechanical system. The entropy  $S$  which is a function of the momentary state of the system is given for the volume element at rest by the expression:<sup>27 · 1</sup>

$$\frac{dS}{dt} = \frac{1}{T} \frac{dQ}{dt} \quad (27 \cdot 5)$$

Considering this definition, eq. 27 · 4 becomes

$$\frac{d\Phi}{dt} = T \frac{dS}{dt} + \frac{dW_i}{dt} \quad (27 \cdot 6)$$

which is the fundamental equation of thermodynamic change of state representing the interrelation of temperature, entropy,

deformation, and stress by means of the specific internal (potential) energy  $\Phi$ ; it is valid for constant mass.

The irreversibility of a change of state requires that

$$\frac{dS}{dt} > 0 \quad (27 \cdot 7)$$

The entropy  $S$  is a function of the same variables as the energy potential  $\Phi$ ; it defines the *thermodynamic state* of a system.

If the system is considered to consist of discrete particles, its *state* is determined by the energy distribution over the particles (see Art. 15). The entropy defining this state must therefore be related in some way with this energy distribution. Since it is evident that a change of state, if it is to occur, will necessarily be accompanied by an increase of the probability associated with the energy distribution over the particles, the entropy, because of its relation to the irreversibility of mechanical processes, can be considered a direct measure  $S = f(P)$  of the *thermodynamical probability*  $P$  of the energy distribution. This is the meaning of the *Boltzmann–Planck* relation  $S = k \log P$ , which establishes the relation between classical and statistical thermodynamics; in this relation the *thermodynamical probability* of a state is defined as the number of equally probable energy distributions over the particles associated with the considered thermodynamical state of the whole system, defined by its entropy.<sup>27·2</sup>

The Boltzmann–Planck relation follows from the definition of the entropy according to which the total entropy of a system is obtained as the sum of the entropies of the constituent parts, since entropy, like energy, is an additive property; the probability of the state of a system, on the other hand, is obtained as the product of the probabilities of the states of the constituent parts. The relation between entropy and probability must therefore be a logarithmic one, since this is the only relation which is suited to express the additive character of entropy in terms of addition of probabilities by multiplication.

Unlike other properties, entropy is a rather abstract concept since it is not directly related to every-day experience. It is a function of the state of a system, and a change of entropy, like one of energy, depends only on the initial and the final states, not on the path between them.

Being a function of the momentary state of the system, and having the tendency to increase with any irreversible change of the system, entropy is in fact the only signpost of time, the only concept by which the passage of time, as expressed by the definitions of *earlier* and *later*, can be defined. In statistical thermodynamics entropy is defined in terms of the distribution of energy over the particles (energy pattern), that is, in terms of the probability of any particular distribution. The actually observed distribution is necessarily associated with a relatively large probability of occurrence; any spontaneous change of state will transfer the system to a state of still higher probability. The system comes to rest when it reaches a state of equilibrium defined by a state of maximum probability. The direction of spontaneous change is thus toward the increasingly probable states of increasing randomness. Changes of entropy can therefore be calculated in terms of the changes of probability, as expressed by the Boltzmann–Planck relation. Such calculations are actually carried out in order to establish the degree of disorder in binary metal crystals and the tendency in the structure of such crystals to change from the randomness of the solid solution to the order of the superlattice (see Art. 11).

It should be pointed out, however, that at normal temperatures most materials exist more or less indefinitely in states that are not those of perfect thermodynamic equilibrium, but rather states in which the tendency to changes has become very small.

According to eq. 27·2 the *power* or time derivative of the work  $A$  of the external (body and surface) forces is

$$\frac{dA}{dt} = \frac{dW_k}{dt} + \frac{d\phi}{dt} \pm \frac{dQ}{dt} \quad (27\cdot8)$$

where the minus sign before  $dQ/dt$  applies for thermal energy added, the plus sign for thermal energy subtracted or *dissipated*.

An *adiabatic* change of state is defined by  $dQ/dt = 0$ . This change is therefore reversible since no energy transformation into heat has taken place. For an isothermal change of state  $dT/dt = 0$ , and eq. 27·8 takes the following form:

$$\frac{dA}{dt} = \frac{dW_k}{dt} + \frac{d}{dt}(\Phi - TS) \quad (27\cdot9)$$

The expression  $(\Phi - TS)$  defines the *free energy* of the system

which, for an isothermal change of state, replaces the energy potential by which the reversible changes of state are governed under adiabatic conditions.

During an irreversible change of state part of the applied external energy is *dissipated* into heat. By denoting this dissipated energy by  $W_D$  and introducing the relation 27·5 in the form,

$$\frac{dW_D}{dt} = - \frac{dQ}{dt} = - T \frac{dS}{dt} \quad (27 \cdot 10)$$

the thermodynamical equation of state for irreversible processes is obtained from eqs. 27·6 and 27·10,

$$\frac{d}{dt}(W\rho) = \frac{d\Phi}{dt} + \frac{dW_D}{dt} - \frac{dW_i}{dt} \quad (27 \cdot 11)$$

where  $W$  denotes the total energy and  $\rho$  the density of the body considered which has to be introduced because the first law applies to unit of mass, whereas the energies  $W_D$ ,  $\Phi$ , and  $W_i$  are usually specified for unit volume. For equilibrium with  $W = \text{const}$ , the time derivative  $\frac{d}{dt}(W\rho) = \frac{d\rho}{dt} W$ ; hence, eq. 27·11 becomes the Gibbs–Helmholtz relation

$$\frac{d\phi}{dt} + \frac{dW_D}{dt} - \frac{dW_i}{dt} - \frac{d\rho}{dt} W = 0 \quad (27 \cdot 12)$$

The power of the work  $W_i$  of the internal forces or stresses or of the work  $A$  of the external forces can always be expressed in terms of the mechanical variables ( $T_{jk}$ ,  $E_{jk}$ ); it is thus mechanically defined. This is not always the case with regard to  $\frac{d}{dt}\Phi$

and  $\frac{d}{dt}W_D$ , which are characteristics of the material, describing respectively, the rate at which the material is able to store applied energy and to dissipate it.

If, in a mechanically closed system, both  $\Phi$  and  $W_D$  are mechanically defined functions, that is, expressible in terms of the mechanical variables, eq. 27·12 is transformed into an equation of state for the material described by the functions  $\Phi$  and  $W_D$ .

According to the second law and the definition of dissipated energy (eq. 27·10), the inequality relation  $\dot{W}_D > 0$  defines irreversible processes; reversibility of the process would be defined by a relation of equality. Since any deformation can be split into its volumetric and distortional components, and since volumetric changes are essentially reversible and have therefore no bearing on the inelastic deformation (see Art. 21), the equation of state governing volume-constant irreversible deformation is obtained from eq. 27·12 by assuming  $\Phi$  and  $W_i$  to refer to distortions only and by introducing  $\dot{\rho} = \dot{e}_v \rho = 0$ :

$$-\frac{dW_{i0}}{dt} + \frac{d\Phi_0}{dt} + \frac{dW_D}{dt} = 0 \quad (27 \cdot 13)$$

where the subscript 0 indicates that the energies are the distortional (deviatoric) components of the respective energies. Because of eq. 27·3, eq. 27·13 is transformed into

$$\frac{dW_{k0}}{dt} + \frac{d\Phi_0}{dt} + \frac{dW_D}{dt} - \frac{dA_0}{dt} = 0 \quad (27 \cdot 14)$$

According to eq. 27·14, the distortional change of state of any deformable body is determined by the momentary distribution of the energy applied per unit time among the three fundamental types of energy: kinetic, potential, and dissipated (heat) energy. The state of the body at any time  $t$  can therefore be expressed in triangular coordinates (Fig. 27·1); this representation was first proposed by Weissenberg.<sup>27·3</sup>

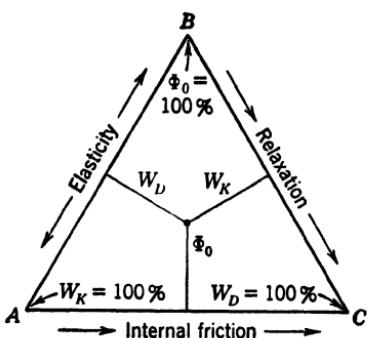


FIG. 27·1 Representation in triangular coordinates of state of energy of body of general behavior. (Arrows indicate possible directions of energy transformation.)

the ideal liquid. Point  $B(A_0 = \Phi_0)$  represents the condition that the applied energy is stored up as potential energy in a stationary state; this is the condition of the perfectly elastic body.

of the energy applied per unit time among the three fundamental types of energy: kinetic, potential, and dissipated (heat) energy. The state of the body at any time  $t$  can therefore be expressed in triangular coordinates (Fig. 27·1); this representation was first proposed by Weissenberg.<sup>27·3</sup> Point  $A$  ( $A_0 = W_{k0}$ ) defines conditions in which the applied energy is immediately converted into kinetic energy. Such conditions are represented by the perfectly rigid solid and

Point  $C(A_0 = W_D)$  describes the immediate dissipation of the applied energy into heat, a condition represented by the perfectly viscous liquid in slow stationary flow, considered as a sequence of states of equilibrium. Line  $AB$  represents the reversible interchange between potential and kinetic energy without loss, as represented by elastic vibrations; line  $BC$  represents the relaxation of potential energy; line  $AC$  represents the dissipation into heat of the kinetic energy either by viscous flow or by solid friction. Energy transformations along  $AC$  and  $BC$  are irreversible in the direction of  $C$ .

The general state of a body is represented by a point in the interior of the triangle. Changes of state are expressed by changes in the triangular coordinates of the point. Thus, the sequence of states during a damped free oscillation may be represented by the diagram shown in Fig. 27·2: the interchange of potential and kinetic energy is associated with energy dissipation; the vibration ceases when all mechanical energy has been transformed (dissipated) into heat energy.

The lines  $AC$  and  $BC$  represent both processes of energy dissipation; however, the points on  $AC$  define processes where dissipation occurs from a position of rest of the body, whereas points on  $BC$  describe dissipation associated with motion. Energy dissipation from a position of rest, in which the applied energy is first stored up as potential energy, is usually defined as *relaxation*. It can only manifest itself in materials which are able to store reversibly a certain energy potential, at least during a finite time. Energy dissipation as a result of motion is defined as *internal friction* in solids, and as *viscosity* in liquids.

The triangular representation of the general energy relation 27·14 indicates that there are three basic types of change of mechanical state of a solid since there are only three types of

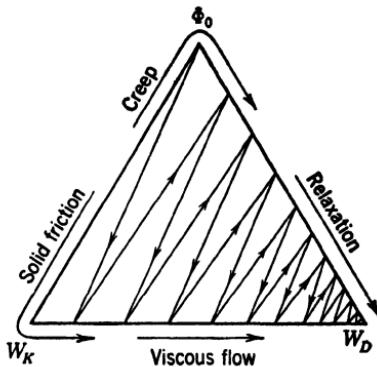


FIG. 27·2 Change of state in terms of energy during damped free oscillation. (Arrows indicate different directions of energy dissipation by solid friction, creep, viscous (fluid) flow, and relaxation.)

energy transformation. These are indicated by the sides of the energy triangle: a *reversible* transformation producing *elasticity*, and two *irreversible* transformations producing, respectively, *relaxation* and *internal friction*. Thus, thermodynamic considerations lead to a subdivision of general mechanical behavior into three classes somewhat similar to those that were derived from considerations concerning the difference between the ordered and the unordered structure of matter (see Art. 4).

Although relaxation and viscosity can be described by the same parameter, the coefficient of viscosity, it is preferable not to identify relaxation with viscosity, but to use *relaxation* in reference to materials in which a *finite* elastic potential can be built up by the applied forces, at least temporarily, and to use *viscosity* in reference to materials for which this is not possible. Thus, the term *relaxation* should be applied to solids, whereas *viscosity* should refer to the behavior of liquids.

If a force is applied to the *relaxing* material, the ensuing deformation process is usually called *creep*. Although this deformation may, in the simplest case, proceed at a linear rate and thus resemble linear viscous flow, there is the difference that creep is the result of progressive relaxation of applied energy by way of the temporary storing up of *finite* amounts of potential energy, whereas in viscous flow the applied energy is directly dissipated by way of kinetic energy; the potentials which are temporarily built up between groups of molecules and rapidly dissipated during flow are *infinitely small*.

It is necessary to distinguish between the ability of a material to store applied energy temporarily and its ability to retain, indefinitely, a limited amount of it. In the first case, the relation between load and deformation or between applied and dissipated energy is continuous over the whole range of deformation. In the second case it must be discontinuous, since the rate of irrecoverable deformation is different for loads associated with energy potentials smaller than the limiting amount of energy that can be stored up indefinitely, and for loads producing energy potentials exceeding this limit. Only materials that possess the ability of reversibly storing up finite amounts of potential energy are real solids; all other apparently solid materials show the behavior of fluids and should be considered as undercooled liquids or "pseudosolids." The limiting energy level represents

the *flow* or *yield limit* in terms of energy; in terms of load or stress it is usually called the *yield value* (in rheology) or the *yield stress* (in engineering).

The existence of a yield limit implies the existence of a *transition point* of mechanical behavior, delimiting conditions under which the applied energy is reversibly stored up from conditions where no further storing up takes place and the energy exceeding the stored-up potential is immediately dissipated. In the triangular energy representation (Fig. 27·1) the finite yield limit defines the condition under which energy transformation starts to take place along the line *BC*. Since some inelastic deformation associated with energy dissipation along the line *AC* will always occur before the applied energy reaches the yield limit, this limit is not an absolute limit of permanent deformation, but an energy (or stress) level at which the mechanism of dissipation undergoes a more or less rapid process of transformation. The comparative magnitudes of the inelastic deformation under loads below the transition level and above this level will frequently justify the assumption that irrecoverable deformation below the yield limit, caused by the relaxation mechanism, is negligibly small and that the yield limit represents a practical limit of permanent deformation. However, in real materials some inelastic deformation is always present even below the yield limit.

Structurally, the discontinuous transition from one mechanism of energy dissipation to another can only be caused by a similarly discontinuous change in the structure of the material. Since real materials are mostly multiple-phase systems, the existence of a yield or flow limit suggests the occurrence of rapid structural changes within at least one phase, if a certain stress level is exceeded. Such changes consist mostly of local breakdowns or of the total destruction of a continuous solid phase in a two-phase solid-fluid system (gels), or of the more rigid phase in a two- or polyphase solid-solid system (polycrystalline metal) or in a polyphase solid-fluid system (high polymer). Single-phase materials, such as glass, have no yield limit. On the other hand, polyphase materials the many phases of which are only moderately different in the grade of their response to external load will have so many consecutive transition points that the concept of a definite yield limit loses its meaning; in this case the points

marking discontinuities in mechanical behavior associated with discontinuous structural changes are becoming so closely spaced as to produce an appearance of continuity.

## 28. Change of State. Energy of the Deformable Body

The forces acting on any volume element produce in it a state of strain and impart to it a certain amount of kinetic energy. The *power*, that is, the time derivative of the energy is the product of the acting forces times the velocity of the resulting displacement and consists of the power of the stresses and the power of the body forces (inertia, gravity). Hence,

$$\dot{A} = s_{11}\dot{e}_{11} + s_{22}\dot{e}_{22} + s_{33}\dot{e}_{33} + s_{12}\dot{g}_{12} + s_{23}\dot{g}_{23} + s_{31}\dot{g}_{31} + \rho(\dot{x}_1\ddot{x}_1 + \dot{x}_2\ddot{x}_2 + \dot{x}_3\ddot{x}_3) \quad (28 \cdot 1)$$

The last term is the time derivative of the kinetic energy  $W_k$ , as can be shown by differentiating with regard to time the expression for  $W_k$  for unit volume,

$$W_k = \frac{1}{2}mv^2 = \frac{\rho}{2}(\dot{x}^2 + \dot{x}_2^2 + \dot{x}_3^2) \quad (28 \cdot 2)$$

The power of the strain energy  $\dot{W}_i$  alone, according to eq. 27·3,  
 $\dot{W}_i = s_{11}\dot{e}_{11} + s_{22}\dot{e}_{22} + s_{33}\dot{e}_{33} + s_{12}\dot{g}_{12} + s_{13}\dot{g}_{13} + s_{23}\dot{g}_{23} \quad (28 \cdot 3)$   
or, in tensor notation,

$$\dot{W}_i = \mathbf{T}_{jk}\dot{\mathbf{E}}_{jk} \quad (28 \cdot 4)$$

The tensor product is defined as the sum of the products of the corresponding components. By introducing the volumetric and deviatoric components of the tensors  $\mathbf{T}_{jk} = \mathbf{T}_v + \mathbf{T}_{0jk}$  and  $\dot{\mathbf{E}}_{jk} = \dot{\mathbf{E}}_v + \dot{\mathbf{E}}_{0jk}$ , and considering that  $\mathbf{T}_{0jk}\dot{\mathbf{E}}_v$  and  $\mathbf{T}_v\dot{\mathbf{E}}_{0jk}$  are necessarily zero, the power of the strain energy can be resolved into the power of the volumetric and of the distortional energies;

$$\dot{W}_i = \mathbf{T}_v\dot{\mathbf{E}}_v + \mathbf{T}_{0jk}\dot{\mathbf{E}}_{0jk} \quad (28 \cdot 5)$$

Whereas  $\mathbf{T}$  and  $\dot{\mathbf{E}}$  are tensors, their products are scalars, the energy being a scalar.

If eq. 28·5 is combined with eq. 27·13, the equation of state for distortion becomes

$$\frac{dW_{i0}}{dt} = \mathbf{T}_{0jk}\dot{\mathbf{E}}_{0jk} = \frac{d}{dt}\Phi_0 + \frac{d}{dt}W_D \quad (28 \cdot 6)$$

where both  $\Phi_0$  and  $W_D$  are functions of the dynamical and kine-

mathematical variables. For volumetric deformation,

$$\mathbf{T}_v \dot{\mathbf{E}}_v = \frac{d}{dt} \Phi_v \quad (28 \cdot 7)$$

The foregoing mechanical equations of state describe in the form of differential equations the relations between the kinematical and the dynamical variables, as defined by their respective tensors. Their constants are the mechanical constants of the material.

If the alternative energy interchanges between two types of energy only are analyzed, while the third vanishes in comparison, eq. 28·6 is split up into three special equations, of which evidently only two are independent, as only two of the three functions appearing in eq. 28·6, namely,  $\Phi_0$  and  $W_D$ , can be chosen independently. Hence,

(a) With  $\dot{W}_D = 0$ ,

$$\dot{W}_{i0} = \frac{d}{dt} \Phi_0(\mathbf{T}_0, \mathbf{E}_0) = \mathbf{T}_{0jk} \dot{\mathbf{E}}_{0jk} \quad (28 \cdot 8)$$

which is the equation of state of the elastic solid. The rate of energy applied equals that of energy stored up; the process is perfectly reversible, since no energy is dissipated. Since  $\Phi_0$  is a function of the time derivatives of  $\mathbf{T}_0$  and  $\mathbf{E}_0$ , respectively, eq. 28·8 defines a general relation of elasticity, of which the linear relation of the classical theory of elasticity represents the simplest special case.

(b) With  $\dot{W}_{i0} = 0$ ,

$$\frac{d}{dt} \Phi_0(\mathbf{T}_0, \mathbf{E}_0) + \frac{d}{dt} W_D(\mathbf{T}_0, \mathbf{E}_0) = 0 \quad (28 \cdot 9)$$

which is an equation describing an interchange between the elastic potential and the dissipated energy under such conditions that no further distortional energy is applied. The process is irreversible and proceeds, according to the second law, in the direction of decreasing potential; it has been designated as relaxation. Thus eq. 28·9 represents the most general law of relaxation.

(c) With  $\Phi_0 = 0$ ,

$$\frac{d}{dt} W_{i0}(\mathbf{T}_0 \mathbf{E}_0) = \frac{d}{dt} W_D(\mathbf{T}_0 \mathbf{E}_0) \quad (28 \cdot 10)$$

which is an equation defining a condition in which the applied strain energy is balanced by the rate of energy dissipated into heat. Processes of this type are completely irreversible; the behavior expressed by eq. 28·10 can be considered the result of internal friction and is usually called plasticity. It represents the most general law of internal friction.

Equations 28·8, 28·9, and 28·10 are the equations of state underlying the theories of elasticity, relaxation and plasticity (internal friction). They have been developed from the basic principles of thermodynamics without any particular assumptions.

## 29. Superposition of Simplified Equations of State

The behavior of real materials can be described by superimposing the three idealized types of behavior specified by eqs. 28·8, 28·9, and 28·10. The material of general deformational behavior is thus represented as a mixture of idealized materials.

In the equation of state 28·6 of a real material it is possible to define the functions,

$$\Phi_0(\mathbf{T}_0, \mathbf{E}_0) = \Phi_{01} \quad (29\cdot1)$$

$$W_n(\mathbf{T}_0, \mathbf{E}_0) = W_{n1}$$

in such a way that two component bodies are described by the relations,

$$\mathbf{T}_{0jk}\dot{\mathbf{E}}_{0jk} = \frac{d}{dt} \Phi_{01} \quad \text{and} \quad \mathbf{T}_{0jk}\dot{\mathbf{E}}_{0jk} = \frac{d}{dt} W_{n1} \quad (29\cdot2)$$

If the general and the two component bodies are subjected to the same stress deviator  $\mathbf{T}_0$ , three solutions exist  $\mathbf{E}_0 = \mathbf{E}_{0k}(\mathbf{T}_0, t)$  where  $k$  takes, consecutively, the values  $k = 1, 2, 3$ . If it is possible to eliminate  $\mathbf{T}_0$  and  $t$  from these three equations, a relation between the three values of  $\dot{\mathbf{E}}_{0k}$  pertaining to the general and the two ideal bodies, respectively, is obtained. By expressing  $\mathbf{E}_{01}$  in terms of  $\mathbf{E}_{02}$  and  $\mathbf{E}_{03}$ , the law of superposition  $\mathbf{E}_{01} = f_1(\mathbf{E}_{02}, \mathbf{E}_{03})$  is established, according to which the deformation of the general material can be computed from those of the component materials.

Similarly, by subjecting the three bodies to the same distortion tensor  $\mathbf{E}_0$ , the solution of the three eqs. 28·6 and 29·2 for  $\mathbf{T}_0$  leads to three equations  $\mathbf{T}_0 = \mathbf{T}_{0k}(\mathbf{E}_0, t)$ ; the subsequent elimina-

tion of  $\mathbf{E}_0$  and  $t$  results in a relation among the three values of  $\mathbf{T}_{0k}$  pertaining to the three bodies. If  $\mathbf{T}_{01}$  is expressed in terms of  $\mathbf{T}_{02}$  and  $\mathbf{T}_{03}$ , the resulting equation is the law of superposition of the stresses  $\mathbf{T}_{01} = f_2(\mathbf{T}_{02}, \mathbf{T}_{03})$ . The functions  $f_1$  and  $f_2$  are the characteristic functions of superposition of equations of state from which the law of superposition of the variables can be derived.

Thus, if the equations of state for the component bodies are given and the behavior of the general body is to be found, all component bodies are subjected to either the same kinematical or the same dynamical conditions. The total dynamical or kinematical response is then obtained by simple addition of the component responses. The establishment of the equation of state of the general body, however, is difficult since the time derivatives of the functions  $\Phi_0$  and  $W_D$  cannot be added. If, on the other hand, the component equations of state of a general body, described by a general equation of state, are to be found, the time derivatives of the constituent functions  $\Phi_0$  and  $W_D$  are usually connected by simple additive relations, whereas the laws of superposition of the mechanical variables are rather complex, so that no direct addition of the variables is possible.

However, a group of general bodies can be defined for which simple additive superposition laws are valid with regard to both the equations of state and the mechanical variables. Only those bodies are actually accessible to mathematical analysis. They are defined by the condition that the functions  $d\Phi_0/dt$  and  $dW_D/dt$  can be expressed in terms of either the dynamical or the kinematical variables alone, or, more generally, that a strain-energy function or a potential  $\Phi_0$  and a dissipation function  $W_D$  exist which are both independent of the sequence of straining. The two types of mechanical equations of state that are obtained by imposing this condition are of the form,

$$\mathbf{T}_0 \dot{\mathbf{E}}_0 = \frac{d}{dt} \Phi_0(\mathbf{E}_0) + \frac{d}{dt} W_D(\mathbf{E}_0) \quad (29 \cdot 3)$$

and

$$\mathbf{T}_0 \dot{\mathbf{E}}_0 = \frac{d}{dt} \Phi_0(\mathbf{T}_0) + \frac{d}{dt} W_D(\mathbf{T}_0) \quad (29 \cdot 4)$$

Since for  $\dot{\mathbf{E}}_0 = 0$  both functions  $\Phi_0(\mathbf{E}_0)$  and  $W_D(\mathbf{E}_0)$  necessarily

vanish, they must be divisible by  $\dot{\mathbf{E}}_0$ . Hence,

$$\mathbf{T}_0 = \frac{1}{\dot{\mathbf{E}}_0} \Phi_0(\mathbf{E}_0) + \frac{1}{\dot{\mathbf{E}}_0} \dot{W}_D(\mathbf{E}_0) = f_1(\mathbf{E}_0, \dot{\mathbf{E}}_0 \dots \mathbf{E}_0^m) \quad (29 \cdot 5)$$

where  $m$  denotes the order of the time derivatives of  $\mathbf{E}_0$ . Similarly, since both functions  $\dot{\Phi}_0(\mathbf{T}_0)$  and  $\dot{W}_D(\mathbf{T}_0)$  vanish for  $\mathbf{T}_0 = 0$ , they must be divisible by  $\mathbf{T}_0$ . Hence,

$$\dot{\mathbf{E}}_0 = \frac{1}{\mathbf{T}_0} \dot{\Phi}_0(\mathbf{T}_0) + \frac{1}{\mathbf{T}_0} \dot{W}_D(\mathbf{T}_0) = f_2(\mathbf{T}_0, \dot{\mathbf{T}}_0 \dots \mathbf{T}_0^n) \quad (29 \cdot 6)$$

where  $n$  denotes the order of the time derivatives of  $\mathbf{T}_0$ .

Under the usually considered conditions of stationary flow all time derivatives of higher than first order can be neglected. The simplified equations of state are therefore

$$\mathbf{T}_0 = f_1(\mathbf{E}_0, \dot{\mathbf{E}}_0) \quad (29 \cdot 7)$$

and

$$\dot{\mathbf{E}}_0 = f_2(\mathbf{T}_0, \dot{\mathbf{T}}_0) \quad (29 \cdot 8)$$

Thus, only in the case of the separation of variables can explicit relations in terms of stresses, strains, and their time derivatives be obtained from the basic form of the equation of state in terms of energies. Equations 29·5 to 29·8 therefore should be always considered as derivations of the fundamental energy relation and not as intrinsic equations of state. Equations 29·7 and 29·8 describe the mathematically simplest types of inelastic behavior. The behavior described by eq. 29·7 is called *after-effect* or *anelasticity* whereas eq. 29·8 describes *relaxation*.

Superposition of equations of after-effect is relatively easy for imposed deformations; by the addition of the stresses produced in the constituent materials the stresses in the combined material are obtained. On the other hand, superposition of equations of relaxation under conditions of imposed stress is performed by simple addition of the strain rates produced by the stress in the constituent materials. The superposition of equations of after-effect under conditions of imposed stress or of equation of relaxation under conditions of imposed deformation is considerably more complex; it requires the integration over time of the

respective equation of state and their solution with regard to either  $E_0$  or  $T_0$ .

By the superposition of a number of idealized materials the behavior of any one of which is described by equations either of after-effect or of relaxation alone, a general behavior is obtained which is no longer one of pure after-effect or of pure relaxation, but a combination of relaxation and after-effect, characteristic of a body described by the general equation of state 28·6.

The simplified equations of state 29·7 and 29·8 have been obtained by introducing the rather arbitrary assumption that the function  $\dot{W}_D$  can be expressed in terms of either the dynamical or the kinematical variables

alone, that is that the dissipation process is independent of sequence. Their validity is thus strictly limited to conditions meeting these assumptions; it should not be expected to extend beyond this range. Observation of the relations  $T_0 = f_1(E_0, \dot{E}_0)$  or  $\dot{E}_0 = f_2(T_0 T_0)$  can therefore be considered observations of equations of state only, if the behavior of the material is simple relaxation or simple after-effect or can be represented by superposition of elements of the same type of simple behavior, differing only in the constants.

Changes of state of any material for which the function  $\dot{W}_D$  cannot be adequately expressed in terms of either the dynamical or the kinematical variables alone (the condition is fulfilled for the elastic potential  $\Phi_0$ ) therefore cannot be expressed in terms of stress, strains, and their time derivatives by equations of the type 29·7 or 29·8, but only in terms of energies, that is, by the general eq. 28·6. There can be no doubt that for the majority of real materials, particularly metals, the dissipation function  $\dot{W}_D$  is not independent of the sequence of deformation, especially beyond the range of very small deformations. Hence, if the conditions under which simplified equations of state can be established are not valid even as a first approximation, equations

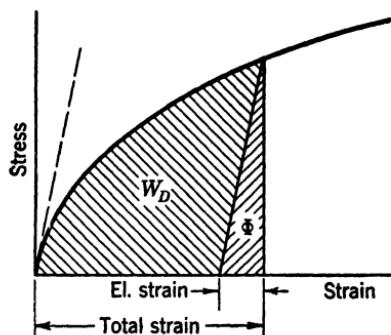


FIG. 29·1 Stress-strain diagram and energies  $W_D$  and  $\Phi$ .

of state should be formulated in terms of energy, by plotting the recoverable (elastic) potential  $\Phi_0$  observed on unloading, as ordinate, against the dissipated energy  $W_D$  associated with the irrecoverable deformation, as abscissa. (Fig. 29·1.) Even this representation will fail, however, to produce an equation of state if part of the dissipated energy is associated with thermal processes taking place during the deformation, which are not or are only indirectly related to the applied load. For three-dimensional states of stress the octahedral shear stress and strain or stress and strain intensities (Art. 41), are to be plotted in Fig. 29·1 since the total distortional strain work,

$$W = \frac{3}{2} \int t_{so} \cdot dg_o = \int s_r de_r \quad (29\cdot 9)$$

### 30. Linear Equations

For all materials the deformational behavior of which can be described by equations of the type 29·5 to 29·8, a relation that is linear to the second approximation may be assumed to exist between the mechanical variables in the vicinity of the origin defined by  $\mathbf{E}_0 = 0$  and  $\mathbf{T}_0 = 0$ , that is, for small strains. This follows from the fact that both the distortional energy potential  $\Phi_0$  and the dissipated energy  $W_D$  must be *even* functions of the strain and its time derivatives, if the direction of the distortion is not to become part of the description of deformational behavior. Such influence of direction would be contrary to the assumption of isotropy of the material. Hence, because of

$$\Phi_0(\mathbf{E}_0^m) = \Phi_0(-\mathbf{E}_0^m) \quad \text{and} \quad W_D(\mathbf{E}_0^m) = W_D(-\mathbf{E}_0^m) \quad (30\cdot 1)$$

the following relations are obtained between the derivatives,

$$\frac{d}{d\mathbf{E}_0} \Phi_0(\mathbf{E}_0^m) = - \frac{d}{d\mathbf{E}_0} \Phi_0^m(\mathbf{E}_0^m) \quad \text{and}$$

$$\frac{d}{d\mathbf{E}_0} W_D(\mathbf{E}_0^m) = - \frac{d}{d\mathbf{E}_0} W_D(-\mathbf{E}_0^m) \quad (30\cdot 2)$$

which indicate that these derivatives are odd function of strain and of its  $m$  time derivatives. Since every odd function has an inflection point at the origin, the functions  $\frac{d}{d\mathbf{E}_0} \Phi_0$  and

$\frac{d}{dE_0} W_D$  can be represented in second approximation by a relation that is linear in  $E_0$  and all its time derivatives. Because of eq. 30·1, all odd terms of the functions  $\Phi_0$  and  $W_D$ , developed into power series, are zero. Since

$$\frac{1}{E_0} \cdot \frac{d\Phi_0}{dt} = \frac{d\Phi_0}{dE_0} \quad \text{and} \quad \frac{1}{\dot{E}_0} \cdot \frac{dW_D}{dt} = \frac{dW_D}{dE_0} \quad (30 \cdot 3)$$

the functions of  $f_1(E_0^n)$  in eqs. 29·5 and 29·7 are linear within the range of small deformations. By a similar reasoning the linearity of the functions  $f_2(T_0^n)$  can be established within the same range. Thus, within the range of small deformations the simplified equations of state of the homogeneous, isotropic body 29·7 and 29·8 are necessarily linear relations between the mechanical variables.

In the mechanical equations of state of deformable materials derived on the basis of thermodynamical considerations, three types of constants can be distinguished: the constants of elasticity, of relaxation, and of internal friction (plasticity). In the particular case of a linear relation between dynamical and kinematical variables, one single constant will define each type of deformational behavior since those constants refer to volume-constant distortions only; an additional constant, however, must be introduced to describe the relation between isotropic stress and volume change, which, for the homogeneous isotropic medium, is an elastic relation (see Art. 27). Hence, under the assumption of linearity of the relations 29·7 and 29·8 four mechanical constants have to be introduced: two constants of elasticity (one referring to volume change and one to distortion), one relaxation constant, and one friction constant; both constants of inelasticity are functions of the absolute temperature  $T$ .

The simplest types of linear deformational behavior can therefore be described by

(a) The equations of linear elasticity:

$$T_0 = 2GE_0 \quad \text{or} \quad \dot{T}_0 = 2G\dot{E}_0 \quad (30 \cdot 4)$$

(b) The equation of linear after-effect:

$$T_0 = 2GE_0 + 2\eta\dot{E}_0 \quad (30 \cdot 5)$$

(c) The equations of linear relaxation:

$$\dot{\mathbf{E}}_0 = \frac{1}{2G} \dot{\mathbf{T}}_0 + \frac{1}{2\eta} \mathbf{T}_0 \quad (30 \cdot 6)$$

In addition to any of the equations of distortion, the volumetric deformation is defined by the relation:

$$\mathbf{T}_v = K \mathbf{E}_v \quad \text{or} \quad \dot{\mathbf{T}}_v = K \dot{\mathbf{E}}_v \quad (30 \cdot 7)$$

In the foregoing equations  $K$  and  $G$  denote, respectively, the *bulk modulus* and the *shear modulus* of elastic deformation and  $\eta$  the coefficients of viscosity.

In all considerations involving the potential  $\Phi_0$  and the dissipation function  $W_D$ , it was tacitly assumed that a function  $\dot{W}_D > 0$  existed for any value of  $\Phi_0$ . It is possible, however, to introduce the assumption that only for values  $\Phi_0 \geq \Phi_y$  the power  $\dot{W}_D$  of the dissipated energy has a finite positive value, whereas for  $\Phi_0 < \Phi_y$  the change of state of the material is perfectly reversible since  $\dot{W}_D = 0$ . This assumption introduces a discontinuity of behavior delimited by an additional constant  $\Phi_y$ , the yield limit, or the yield stress  $s_0$  (see Art. 27).

The linear inelastic behavior described by the constants  $K$ ,  $G$ ,  $\eta$  and  $\Phi_y$  or  $s_0$  is the simplest approximation of real behavior. For many materials this simple approximation is not sufficient, and the "constants" become dependent on stress. This is an expression of the fact that they vary with the internal structure, which is permanently changed in the course of the inelastic deformation. An empirical expression for the "constant" as a function of stress may in such cases be established on the basis of test results, and expressed by a power series in stress,

$$e = \alpha_1 s + \alpha_3 s^3 + \alpha_5 s^5 + \dots \quad (30 \cdot 8)$$

where  $\alpha_1$  represents the modulus of linear elasticity.

A more frequently used method of reproducing nonlinear behavior is the use of power functions with nonintegral powers. In this case the relation of stress and strain is usually written in one of the forms,

$$e = \alpha s^m \quad \text{or} \quad s = \beta e^n \quad (30 \cdot 9)$$

where the exponent  $m > 1$  or  $n < 1$ .

The use of power laws, although convenient, particularly

because of the linearity in double logarithmic representation, has however, definite limitations.<sup>30·1</sup> The first is in the fact that for very small values of either variable the behavior of the power function becomes irregular and differs appreciably from its behavior at larger values, whereas the representation of experimental results usually requires that for small values of the variables the relations become linear. For  $e = 0$  the tangent modulus of the power law will be either infinite or zero, depending on whether the power  $\geq 1$ . Power functions must therefore not be extrapolated to zero.

The second limitation is that of variability of dimensions. Since  $e$  and  $s$  have definite dimensions and the exponent  $m$  must evidently be dimensionless, the dimension of  $\alpha$  or  $\beta$ , respectively, will vary with each value of the exponent. The law describes, therefore, particular conditions, not the general behavior of the material. There are, however, two conditions under which a power law may be used to express a general functional relation without the dimensional objection, namely: if the exponent has a constant value throughout, or if the variables are dimensionless, being introduced as ratios  $(s/s_0)$  and  $(e/e_0)$ .

The third limitation is that both variables increase indefinitely, so that for infinite strain the stress becomes infinite. This behavior has no physical parallel.

The principle advantage in the use of power laws is that, in double logarithmic representation, all power relations are straight lines; thus,

$$\log e = \log \alpha + m \log s \quad \text{or} \quad \log s = \log \beta + n \log e \quad (30 \cdot 10)$$

The slope of this relation is obtained by differentiation of eq. 30·10:

$$\frac{d \log s}{d \log e} = \frac{1}{m} \quad \text{or} \quad \frac{d \log s}{d \log e} = n \quad (30 \cdot 11)$$

Thus for power laws in logarithmic representation the slope of the diagram is the inverse of the exponent or the exponent itself, depending on which of the equations 30·9 is used. There is no physical significance in this fact.

Power laws are therefore nothing but simple interpolation formulas; they should be used only within the range in which their validity has been established.

### 31. Applications of the Concept of the Equation of State

The concept of a mechanical equation of state applicable in the interpretation of mechanical tests was first suggested by Ludwik.<sup>31·1</sup> However, practical applications of this concept by which stress, strain, strain rate, and temperature can be interrelated have only been made rather recently in experiments with metals, particularly in connection with the aim of studying the interrelation between tensile tests and creep tests.<sup>31·2</sup> The establishment of a valid equation of state, for instance, would eliminate the necessity for long-time creep tests; in general, by the aid of such an equation the expected behavior under difficult testing conditions could be derived from the results of easily performable tests.

Since the existence of a mechanical equation of state implies that the stress necessary to produce a given strain rate at a given temperature depends only on the instantaneous values of temperature and strain, no structural change must be produced in the course of the deformation that would make the final structure of the material identifiably different from its initial structure. If the change is identified by comparative values of stress at given strain, strain rate, and temperature, even relatively small structural changes may invalidate the concept of an equation of state. However, if the identification is in terms of changes of the potential energy  $\Phi$  at a given value  $W_D$ , an equation of state (expressed in energies) will be more likely to exist, in spite of a certain amount of structural change, since a considerable number of states of a material body which are different when defined in terms of the mechanical variables and their time derivatives may prove to be very nearly identical if expressed in terms of energy.

Structural changes, which are functionally unrelated to the mechanically applied energy, must certainly be expected to invalidate the equation of state, unless such changes, although unrelated to the mechanical energy, could be quantitatively related to changes of entropy. Some attempts in this direction have recently been made.<sup>31·3</sup>

At temperatures sufficiently below recrystallization temperature, metals do not undergo "spontaneous," that is, thermal, structural changes during moderate time intervals; similarly, for small amounts of inelastic strain, recrystallization proceeds

very slowly. Thus, the concept of a continuous equation of state is frequently valid as a first approximation, particularly if the considered range of temperature variation is moderate.

If isothermal tensile tests at different constant strain rates are performed under conditions for which an equation of state exists, a family of stress-strain curves is obtained; similarly another family of curves may be obtained by performing a number of tensile tests at constant strain rate and various temperatures (Fig. 31·1). If in the first series of tests one test is interrupted at a certain stress, the straining necessarily proceeds under constant stress at a decreasing rate. If the previous strain rate is suddenly reapplied, the stress is raised to the level it would have attained at the same strain in the uninterrupted test. Similarly, if the temperature is suddenly changed

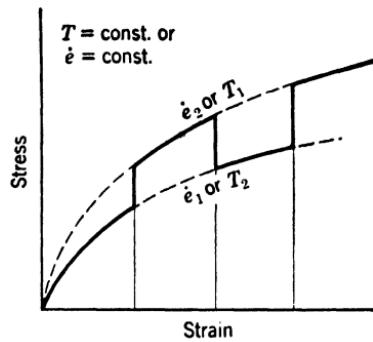


FIG. 31·1 Verification of a mechanical equation of state.

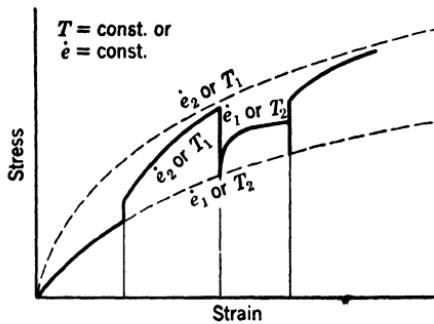


FIG. 31·2 Failure of a mechanical equation of state.

during a test at constant strain rate, the stress level follows this change of temperature, and the stress-strain relation assumes the form pertaining to the new temperature of the test, and continues as if the whole test had been performed at the new temperature. Investigations conducted on a number of metals and alloys<sup>31·4</sup> with the purpose of checking the existence of a mechanical

equation of state over a considerable range of strain have shown that the stress at a certain strain does not depend on the momentary value of strain rate and temperature alone, but also on the temperature at which the specimen was strained previously, since this temperature exerts a marked effect on the structural changes associated with the strain. Thus, the observed curves

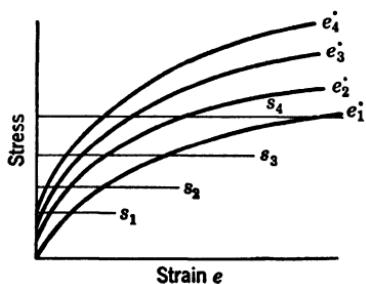


FIG. 31·3 Relations  $s = f_1(e)$  for varying parameter  $\dot{e}$ .

stress and strain are found to provide fairly good evidence for its validity.<sup>31·5</sup>

In order to combine the results of tests at constant strain rate and tests at constant temperature a number of values of the strain-rate-temperature parameter coordinates ( $s, e$ ) can be computed (see Art. 19) and plotted in the stress-strain plane. By suitably interpolating continuous lines of equal value of the parameter  $P = \text{const}$ , equations of state of the form  $s = s(e, P)$  could be established, on the basis of which the combined effect of changes of temperature and of strain rate might be predicted.

The principal practical use of equations of state is in converting data obtained from one type of test into data that might be expected to result from a different type of test. If, from the results of a number of tensile tests at various strain rates  $\dot{e}$ , which have been plotted schematically in Fig. 31·3, the values of the strains  $e$  at different strain rates pertaining to a constant

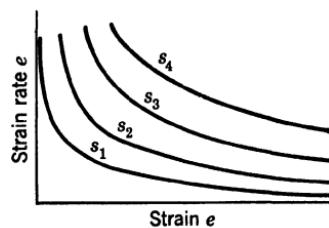


FIG. 31·4 Relations  $\dot{e} = f_2(e)$  for varying parameter  $s$ .

level of stress are derived for a number of stress levels as the points of intersection of the horizontal  $s = \text{const}$  with the curves  $s = f_1(e)_{(e=\text{const})}$  the relations  $\dot{e} = f_2(e)_{(s=\text{const})}$  may be plotted (Fig. 31·4) as well as the relation  $1/\dot{e} = f_3(e)_{(s=\text{const})}$ . By integrating the relation,

$$\frac{de}{dt} = \frac{1}{f_3(e)} \quad (31 \cdot 1)$$

and solving the resulting equation,

$$\int f_3(e) de = t + \text{const} \quad (31 \cdot 2)$$

with regard to  $e$ , the creep function  $e = f_4(t)_{(s=\text{const})}$  is obtained.

On the other hand, if creep curves are recorded at various temperatures (Fig. 31·5), the creep rates as function of the temperature at various given values of elongation may be obtained at the points of intersection of various lines  $e = \text{const}$  with the observed curves  $(e, t)_{(T=\text{const})}$ . In the case of viscous creep, for instance, the functions obtained, according to eq. 19·8, should be of the type:

$$[\text{const} - \ln(\dot{e})]_{(e=\text{const})} = \text{const} \left( \frac{1}{T} \right) \quad (31 \cdot 3)$$

Such relations have actually been observed.<sup>31·6</sup>

### 32. Limit of Continuous Change of State. Fracture

The change of state proceeding continuously within the system of coordinates of the mechanical variables reaches a limit at the point of fracture. The critical energy potential  $\Phi_s$  defining macroscopic fracture (see Art. 22) depends on both initiation and progress of separation on the atomic level. It is therefore essentially dependent on the structural pattern of the material and its change during deformation. Changes in pattern, proceeding under load or spontaneously, will necessarily produce variations in the critical energy potential.

Fracture in two- or polyphase materials must depend on the interaction between phases during the preceding change of state and on the relative importance of the constituent phases in

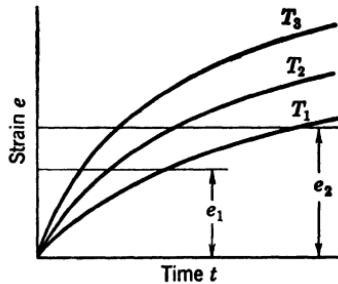


FIG. 31·5 Creep curves at different temperatures  $T$ .

producing changes of structural pattern. Thus, under conditions in which in a polycrystalline aggregate the energy transformation associated with changes of structural pattern is essentially concentrated within the intercrystalline boundaries, the value of the separation potential  $\Phi_s$  will depend predominantly on changes within the intercrystalline phase, in which separation will also take place. Under different conditions the changes of pattern will mainly affect the crystalline regions; the variation

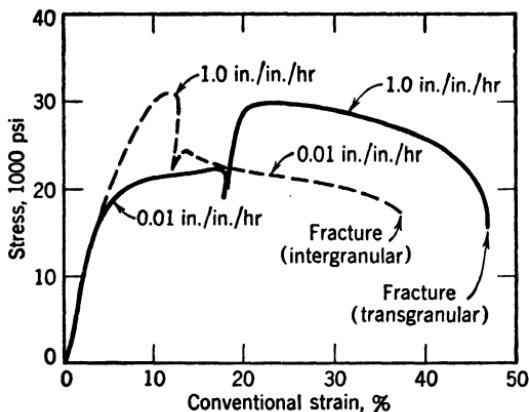


FIG. 32·1 Effect of change of strain rate on deformation and rupture of carbon-molybdenum steel at 1000°F (After R. F. Miller, V. G. Smith, and G. L. Kehl<sup>32·2</sup>).

in the level of the critical energy  $\Phi_s$  will then be essentially related to changes within these regions, and separation must be expected to proceed mainly through the crystals. Since the conditions defining the character of the process of change of state are strain rate and temperature, the resulting type of fracture will necessarily depend on both of these parameters in combination, in the same way in which they determine the change of state immediately preceding fracture. If changes of pattern are produced simultaneously in both phases, the expected character of separation can no longer be predicted and may be of either type. Such conditions are present in polycrystalline metal aggregates near recrystallization temperature, where the thermal stability of particles in both phases is nearly equal and place change of particles between the phases is going on. It has been observed that at such temperatures, called *equicohesive tem-*

peratures, no definite character of fracture develops, intercrystalline separation being as frequent as transcrystalline separation.<sup>32·1</sup>

The stress-strain diagrams reproduced in Fig. 32·1 illustrate the interrelation between the character of fracture and preceding deformation under conditions in which an equation of state can be assumed to exist; they show that the character of fracture does not depend to any appreciable degree on previous history of straining, but only on the momentary values of the strain rate.<sup>32·2</sup>

A more extensive discussion of the interrelation between the structural and the phenomenological aspect of fracture is presented in Chapter 12.

### References

- 27·1 J. C. SLATER and N. H. FRANK, *Introduction to Theoretical Physics*, McGraw-Hill Book Co., New York (1933) 460.
- 27·2 M. BORN, *Atomic Physics*, Blackie & Son, London (1945) 350.
- 27·3 R. EISENSCHITZ, B. RABINOWITSCH, and K. WEISSENBERG, *Mitt. deut. Materialprüfungsanstalt*. **9** (1929) 91.
- 30·1 M. REINER, *Naturwissenschaften* **21** (1933) 294.
- 31·1 P. LUDWIK, *Elemente d. technologischen Mechanik*, J. Springer, Berlin (1909) 44.
- 31·2 C. ZENER and J. H. HOLLOWAY, *J. Applied Phys.* **17** (1946) 69.  
J. H. HOLLOWAY, *Trans. AIME* **162** (1945) 268; **171** (1947) 535.
- 31·3 J. C. FISHER and C. W. MACGREGOR, *Trans. Am. Soc. Metals* **40** (1948) 302.
- 31·4 J. E. DORN, A. GOLDBERG, and T. E. TIETZ, *Metals Technol.* (1948) T.P. 2445.  
E. OROWAN, *The Creep of Metals*, West Coast of Scotland Iron & Steel Inst. (1947) 71.
- 31·5 R. L. FULLMAN and D. TURNBULL, *J. Metals* **1** (1949) 102.
- 31·6 S. DUSHMAN, L. W. DUNBAR, and H. HUTHSTEINER, *J. Applied Phys.* **15** (1944) 108.
- 32·1 W. SIEGFRIED, *J. Applied Mechanics* **10** (1943) A-202.
- 32·2 R. F. MILLER, V. G. SMITH, and G. L. KEHL, *Trans. Am. Soc. Metals* **31** (1943) 817.

## LINEAR BEHAVIOR

**33. Elasticity. Small Strain**

A material is truly elastic if the elastic potential is associated with the forces of the primary (atomic) bonds. The elasticity may be expected to be the more perfect, the more nearly equal the distribution of bond energies over the particles.

The truly elastic deformations of materials are relatively small; they are reversible, without observable energy loss. The relations of force and displacement or of stress and strain are practically linear, as must be expected on the basis of the discussion in Art. 30. Only if the deformations are no longer small, is it necessary to distinguish between real time-independent nonlinear elasticity and the nonlinearity resulting from the existence of time effects. These latter effects produce nonlinear stress-strain relations in loading, accompanied by complete or partial reversal of deformation in unloading. However, even if the reversal is complete, the loading and unloading diagrams are not identical, and energy is dissipated during each loading cycle. This is actually the behavior of most of the materials considered in engineering to manifest nonlinear elasticity. In fact, true nonlinear elasticity cannot be a property of real materials, except for large deformations.

It is generally assumed that the classical theory of elasticity involves two approximations in addition to the assumptions of homogeneity and isotropy: These are the assumptions of infinitesimal deformations and of linearity of the relation between the tensors of stress and of strain. From the discussion of the

character of the functions  $\Phi_0(E_0)$  and  $W_D(E_0)$  in Art. 30, it is evident that one of these two assumptions is redundant, since for an isotropic body the condition of linearity follows directly from the assumption of small deformations and within this range therefore need not be specified as an empirically established additional relation. Since the equations of the classical theory of elasticity are linear, superposition of solutions is permissible. The general equation of the elastic body 28·8, derived from thermodynamical considerations, shows that the behavior of this body is fully defined by the elastic potential or the "strain-energy function"  $\Phi$ , made up of  $\Phi_0$  and  $\Phi_v$ .

As the strain-energy function of the isotropic body is necessarily invariant against rotation of the coordinate axes, it must be a function of the invariants of strain and of temperature or entropy; hence,

$$\Phi = \Phi(I_{e1}, I_{e2}, I_{e3}, T) \quad (33\cdot1)$$

Hypotheses as to the actual form of the strain-energy function must ultimately be justified by experiment. If it is assumed that  $\Phi$  is an analytic function of the strain invariants  $I_{ei} = I_i$ , the elastic potential for constant temperature can be represented by the power series in  $e$ :

$$\Phi = aI_1 + bI_1^2 + cI_2 + dI_1^3 + eI_1I_2 + fI_3 + \dots \quad (33\cdot2)$$

where the coefficients  $a, b, c, \dots$  are parameters that depend on the elastic properties of the medium. The omission of the constant terms from this series indicates the choice of the zero value of  $\Phi$  as the origin; if the stress is zero in the unstrained state, then  $a = 0$ , and the first term involving  $I_1$  vanishes. It has been shown in Art. 30 that congruence of the relation between the mechanical variables in volume-constant tension and compression is possible only if  $\Phi$  is an even function of strain. Hence, eq. 33·2 retains but the terms which are even in strain and may therefore be written in the form:

$$\Phi = bI_1^2 + cI_2 + gI_1^4 + hI_1^2I_2 + jI_2^2 + kI_1I_3 + \dots \quad (33\cdot3)$$

The strain-energy function of (volume-constant) distortion is therefore

$$\Phi_0 = cI_2 + jI_2^2 + \dots \quad (33\cdot4)$$

since for small deformation  $3e_v = I_1 = 0$ . Hence, the energy function of volume change,

$$\Phi_v = bI_1^2 + gI_1^4 + hI_1^2I_2 + kI_1I_3 + \dots \quad (33 \cdot 5)$$

Equations of similar type can be developed in terms of the invariants of stress.

The linear stress-strain relations of the classical theory are obtained by neglecting in eq. 33.3 all terms of higher than second-order in strains. The strain energy function of linear elasticity thus appears as a homogeneous quadratic function of the components of strain:

$$\Phi = bI_1^2 + cI_2 \quad \text{and} \quad \Phi_0 = cI_2 \quad (33 \cdot 6)$$

The coefficients  $b$  and  $c$  are usually expressed in terms of the bulk modulus  $K$ , the modulus of rigidity  $G$ , and Poisson's ratio  $\mu$ :

$$b = \frac{1 - \mu}{1 - 2\mu} G = \frac{3}{2} K \frac{1 - \mu}{1 + \mu} \quad \text{and} \quad c = -2G \quad (33 \cdot 7)$$

Hence,

$$\Phi = \frac{1}{2}KI_1^2 - 2GI_{02} \quad (33 \cdot 8)$$

If the change from the initial to the stressed state is adiabatic (constant entropy  $S$ ) the strain-energy function of the classical theory has the property that

$$\frac{\partial \Phi}{\partial e_i} = s_i \quad \text{and} \quad \frac{\partial \Phi}{\partial s_i} = e_i \quad (33 \cdot 9)$$

where  $e_i$  and  $s_i$  are the components of the strain and the stress tensor, respectively. In isothermal processes the increase of entropy (loss of heat to the surrounding medium) accompanying the change from the initial to the final state must be introduced; hence, the expression of "free energy" ( $\Phi - TS$ ) given in eq. 27.9 plays the role of the energy potential. Thus,

$$\frac{\partial(\Phi - TS)}{\partial e_i} = s_i \quad \text{and} \quad \frac{\partial(\Phi - TS)}{\partial s_i} = e_i \quad (33 \cdot 10)$$

The difference between adiabatic and isothermal elastic deformation which appears in the strain-energy function necessarily affects the values of the elastic constants. If during deformation

the volume of the body is changed, a change in temperature results, as defined by the equation of state for volume changes governing any isotropic material, according to which expansion is accompanied by a drop of temperature, compression by a temperature increase. In adiabatic processes, therefore, the kinematical variables contain a component resulting from changes in dimensions due to the temperature change, which is absent in isothermal processes. This component reduces the amount of deformation if the temperature decreases and increases it if the temperature is raised. Thus, the apparent bulk modulus of adiabatic deformation is higher than that of isothermal deformation if the deformation is accompanied by volume expansion, and lower if it is accompanied by volume compression. Only under conditions of volume-constant distortion (shear), does no difference exist between the adiabatic and isothermal moduli. The numerical differences for metals at room temperature are of the order of magnitude of 1 percent.<sup>33·1</sup>

Attempts have been made to develop a theory of nonlinear elasticity, retaining the assumption of infinitesimal deformations, but introducing even and odd terms of  $e$ , up to the third order.<sup>33·2</sup> The number of independent elastic parameters is thus extended to five, as may be seen from eq. 33·2. Such a potential however does not lead to congruence of the stress-strain relation in tension and in compressions. True nonlinear elasticity is therefore probably not a property associated with any real material under conditions of small strains.

Important problems of apparently nonlinear truly elastic behavior are those involving infinitesimal strains accompanied by large rotations. Although the assumption of linearity of the relation between stress and strain remains valid, the linearity of the relations between forces and deformations no longer exists, because of the effect of the finite rotations.<sup>33·3</sup>

### 34. High Elasticity, Finite Strain

For the manifestation of *high* elasticity of a material, completely different conditions are required from those for *true* elasticity. These conditions are, essentially, the existence of a skeleton structure, that is, of a continuous solid network, in which bonds of widely different energy contents, both primary and secondary, are present. Such networks which are usually

immersed in or filled with a fluid phase are mostly formed by long chain molecules that are either flexible or rigidly inter-linked. If the molecules are short and rigid the network may still be flexible, as the elastic flexibility of a skeleton or of an open framework is large in spite of the rigidity of the connections and of the individual members.<sup>34-1</sup> If the chains forming the network are very flexible, the concept of thermal oscillations of molecules must be introduced to explain the very high elasticity present, as for instance in rubber and rubber-like compounds.

These so-called thermodynamical (kinetic) theories<sup>34-2</sup> of elasticity assume that a chain molecule in the unstretched state will take the statistically most probable form and therefore will not be straight, since the straight line, being one particular form of many alternatives, is rather improbable. If the chains are stretched and approach straight lines, the probability of this stretched state is evidently smaller than that of the unstretched state. Hence, the material tends to re-establish the more probable forms of the molecules in the unstretched state, and this tendency produces the reversibility of the deformation. Kinetic theories assume the molecules to be in the form not of simple chains but of coiled spirals, interlinked by secondary bonds; under load these spirals are uncoiled and stretched but only after the secondary bonds have been disrupted. Since thermal oscillations of particles forming the chain molecules take place essentially in directions perpendicular to the axis of the chains (see Art. 10), the *orientation* of the molecular chains or coils by stretching in the course of which the chains become parallel, approaching each other, produces an increased repulsion between the constituent molecules of neighboring chains; this repulsion which is the result of the intensified thermal oscillations tends to increase the lateral distances between the interacting chains, thus producing lateral expansion associated with longitudinal contraction.

The assumption of lateral intermolecular repulsion explains the anomalous *thermoelastic* or *Joule effect* of rubber as well as the abnormal values of  $\mu > \frac{1}{2}$ . Whereas elastically normal materials show a cooling effect when stretched rather rapidly (adiabatically), rubber exhibits the opposite behavior, developing heat. This heat effect is so pronounced that it can be easily observed by stretching a rubber band quickly and putting it to

the lips; conversely, stretched rubber contracts longitudinally when heated. These effects are assumed to be the result of the intensified lateral thermal oscillations due to stretching or to heating: The resulting lateral expansion produces longitudinal contraction; conversely, heat energy is developed as a result of the oscillations intensified by the crowding together during the stretching of the molecular chains. These anomalous effects, however, accompany large strains only; both the thermo-elastic effect and Poisson's ratio of rubber are normal for small elongations.

The stress-strain relations of high elasticity are not linear. For a large number of materials they are even not monotonous. Thus, although for rubber-like materials the range of high elasticity starts practically at zero stress, a short range of true elasticity precedes the range of high elasticity in films and fibers. The boundary between the two ranges is attained when the secondary cross links between the chains are disrupted and the character of the deformation changes from the elastic straining of the cross-linked structure to the uncoiling and stretching of the unlinked chain molecules; this boundary represents a true yield point of the material. Since with progressing uncoiling and stretching of the molecules the resistance to deformation will necessarily increase, until further deformation can only proceed by elastic stretching of the straightened parallel chains, highly elastic materials appear to "stiffen" with increasing deformation; for large strains their stress-strain diagrams thus tend to bend away from the strain axis (Fig. 20·1).

The development of a theory of finite elastic strain has not made any appreciable progress within the last few decades. For states of pure strain the logarithmic definition of strain extends the validity of the procedure of simple superposition to finite strains. For such states Hencky has attempted to derive stress-strain relations from a second-order strain-energy function.<sup>34·3</sup> The developed equations indicate a volume dependence of both the shear and the bulk modulus for finite strain. This apparent increase of the shear modulus and of the bulk modulus has actually been observed for materials subjected to very high hydrostatic pressures.<sup>34·4</sup>

Attempts have been made to derive stress-strain relations for finite strains in highly elastic materials from rather arbitrarily

assumed analytical strain-energy functions containing more than the two constants of the classical function.<sup>34·5</sup> A theory of finite strains including rotation has first been developed by Hamel,<sup>34·6</sup> and restated and applied to special problems by Murnaghan and Seth;<sup>34·7</sup> a general formulation is due to Reiner.<sup>34·8</sup> According to the theory of finite strains the basic assumption of the classical theory of elasticity, that a distortion is uniquely related to a deviatoric stress and a volume change to an isotropic stress, is valid only within the range of infinitesimal strains. When the strains are finite, a hydrostatic stress will cause a volume change, and vice versa; however such a volume change may also be caused by a deviatoric stress in the absence of a hydrostatic stress. On the other hand, a hydrostatic stress may be required to maintain a certain distortion. It follows from the analysis of finite shear<sup>34·9</sup> that a simple shear is accompanied by a volume expansion; hence, a bar subjected to a torsional moment will be extended in the direction of its axis. This conclusion is borne out by experiments.

The elastic body under conditions of finite strain has therefore a property that is absent in the classical body and that is called *dilatancy*. The existence of this property of volume change under a shearing stress has already been expected by Lord Kelvin,<sup>34·10</sup> who also predicted that the volume change would be proportional to the square of the shear by which it is produced; this assumption is confirmed by the analysis of the elastic body subject to finite strain.

### 35. Linear Bodies. Systematics of Inelastic Behavior

INTEGRATION OF BASIC EQUATIONS. Under the assumption that both the elastic potential  $\Phi$  and the dissipation function  $W_D$  are functions of second order in terms of stresses or strains, the classification of deformable bodies presented in Table 35·1 gives a systematic survey of linear inelastic behavior. Two equations are necessary to describe the total deformation, one for volume change (which is invariably a relation of elasticity) and the second for distortion. Table 35·1 illustrates the formation of general linear inelastic behavior as a synthesis of the behavior of solid and liquid phases, the limiting cases being that of the perfectly elastic solid and of the ideal viscous liquid. In order to facilitate identification a name has been attributed to every

one of the simple materials, the name being that of the first proposer of the respective equation of idealized behavior; this procedure has been suggested by Reiner.<sup>35.1</sup> Thus, the linear elastic solid is termed *Hookean*, whereas the ideal viscous liquid is called *Newtonian*; the linear relaxing material is named after *Maxwell*, the linear anelastic body (after-effect) is named after *Kelvin*, the perfectly plastic solid, the deformation of which is elastic below the yield point and plastic above it, after *St. Venant*, and the general elastic body after *Hencky*. The visco-plastic or *Bingham* body, which has not been included in the table but is frequently mentioned in rheological literature, is actually a generalized St. Venant body whose deformation beyond the yield limit is viscous.

The materials defined in Table 35.1 by the simplified linear equations of state are the *classical bodies*. Their behavior in terms of the variables which are observable in simple mechanical tests is obtained by integrating the respective equation for the volume element, considering the boundary and loading conditions of the test.

The solution of the equations of the Kelvin and of the Maxwell body are obtained as solutions of linear differential equations of the form,

$$\frac{dy}{dx} + My = N \quad (35.1)$$

given by the general expression

$$y = e^{-\int M dx} \cdot (\int Ne^{\int M dx} dt + C) \quad (35.2)$$

Hence, for the Kelvin body with  $M = G/\eta$ ,  $N = \frac{1}{2\eta} \mathbf{T}_0$ , and

$y = \mathbf{E}_0$ :

$$\mathbf{E}_0 = e^{(-G/\eta)t} \cdot \left( \mathbf{E}_{00} + \frac{1}{2\eta} \int_0^t \mathbf{T}_0 e^{(G/\eta)t} dt \right) \quad (35.3)$$

where  $\mathbf{E}_{00}$  denotes the value of  $\mathbf{E}_0$  at  $t = 0$ .

For the Maxwell body with  $M = G/\eta$ ,  $N = 2G\dot{\mathbf{E}}_0$ , and  $y = \mathbf{T}_0$ :

$$\mathbf{T}_0 = e^{(-G/\eta)t} \cdot \left( \mathbf{T}_{00} + 2G \int_0^t \dot{\mathbf{E}}_0 e^{(G/\eta)t} dt \right) \quad (35.4)$$

where  $\mathbf{T}_{00}$  denotes the value of  $\mathbf{T}_0$  at  $t = 0$ .

TABLE 35·1  
LINEAR INELASTIC MATERIALS

Phase	Solid			Liquid		
	Elastic	Anelastic (Aftereffect)	Plastic	Complex (Inelastic)	Relaxing	
Deformation	Hooke	Hencky	Kelvin	St. Venant	General Inelastic Bodies	
Name				Maxwell	Newton	
Equation	$\mathbf{T}_v = K\mathbf{E}_v$	$\dot{\mathbf{T}}_v = K\dot{\mathbf{E}}_v$	$\mathbf{T}_v = K\mathbf{E}_v$	$\dot{\mathbf{T}}_v = K\dot{\mathbf{E}}_v$	$\ddot{\mathbf{T}}_v = K\ddot{\mathbf{E}}_v$	
Volume change	$\mathbf{T}_v = K\mathbf{E}_v$	$\dot{\mathbf{T}}_v = K\dot{\mathbf{E}}_v$	$\mathbf{T}_v = K\mathbf{E}_v$	$\dot{\mathbf{T}}_v = K\dot{\mathbf{E}}_v$	$\ddot{\mathbf{T}}_v = K\ddot{\mathbf{E}}_v$	
Distortion	$\mathbf{T}_0 = 2G\mathbf{E}_0$	$\dot{\mathbf{T}}_0 = 2G\dot{\mathbf{E}}_0$	$\mathbf{T}_0 = 2G\mathbf{E}_0 + 2\eta\dot{\mathbf{E}}_0$	$\frac{\mathbf{T}_0 = 2G\mathbf{E}_0^*}{\mathbf{T}_0 = f(\mathbf{Y}_0)}$ $f_1(\mathbf{T}_0\mathbf{E}_0\dot{\mathbf{E}}_0) + f_2(\mathbf{T}_0\dot{\mathbf{E}}_0\dot{\mathbf{E}}_0) = 0$	$\dot{\mathbf{E}}_0 = \frac{1}{2G}\dot{\mathbf{T}}_0 + \frac{1}{2\eta}\mathbf{T}_0$	$\mathbf{T}_0 = 2\eta\dot{\mathbf{E}}_0$
Constants	$K, G$	$K, G$	$K, G, \eta; \tau = \eta/G$	$K, G, Y_0$	$K, G, \eta; \tau = \eta/G$	$K, \eta$

\* For  $\mathbf{T}_0 < f(\mathbf{Y}_0)$ .

If the elastic solid is subjected to a constant stress deviator  $\mathbf{T}_0 = \mathbf{T}_{00}$ , the resultant strain  $\mathbf{E}_0 = \frac{1}{2G} \mathbf{T}_{00}$  is reached instantaneously. For the Kelvin body the solution for an imposed constant stress deviator  $\mathbf{T}_0 = \mathbf{T}_{00}$  is obtained from eq. 35·5:

$$\mathbf{E}_0 = e^{-(G/\eta)t} \cdot \mathbf{E}_{00} + \frac{1}{2G} \mathbf{T}_{00}(1 - e^{-(G/\eta)t}) \quad (35\cdot5)$$

If, for  $t = 0$ ,  $\mathbf{E}_{00} = 0$ , the time-deformation curve:

$$\mathbf{E}_0 = \frac{1}{2G} \mathbf{T}_{00}(1 - e^{-(G/\eta)t}) \quad (35\cdot6)$$

The final total deformation (for  $t = \infty$ ), which is that of the perfectly elastic solid, is reached asymptotically.

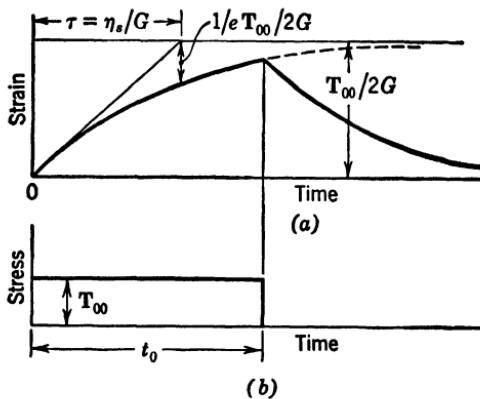


FIG. 35·1 Strain-time curve of Kelvin body for given stress-time diagram.

If at any time  $t_0$  the stress is released, the strain is not immediately recovered; after load release it decreases according to the relation:

$$\mathbf{E}_0 = \frac{1}{2G} (1 - e^{-(G/\eta)t_0}) \cdot e^{-(G/\eta)(t-t_0)} \quad \text{for } t > t_0 \quad (35\cdot7)$$

Equation 35·7 describes the aftereffect of the Kelvin body (Fig. 35·1).

The ratio  $\tau = \eta/G$  has the dimension of time; it is called the *retardation time*. The tangent to the strain-time diagram 35·6:

$$\frac{d\mathbf{E}_0}{dt} = \frac{\mathbf{T}_{00}}{2G} \cdot \frac{G}{\eta} e^{-(G/\eta)t} = \frac{\mathbf{T}_{00}}{2G} \cdot \frac{1}{\tau} e^{-t/\tau} \quad (35 \cdot 8)$$

For  $t = 0$ :

$$\left( \frac{d\mathbf{E}_0}{dt} \right)_{(t=0)} = \frac{\mathbf{T}_{00}}{2G} \cdot \frac{1}{\tau} \quad (35 \cdot 9)$$

Hence, the retardation time is cut off on the horizontal asymptote  $\mathbf{E}_0 = \mathbf{T}_{00}/2G$  of the strain-time function by the tangent to this curve at the origin. It is the time required to produce  $(1 - 1/e)$  of the full elastic deformation under an applied constant stress (Fig. 35·1).

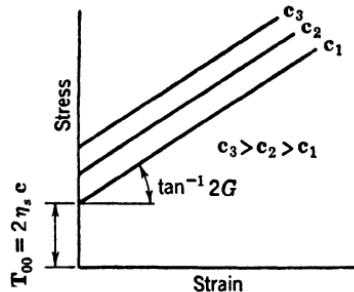


FIG. 35·2 Stress-strain relation for Kelvin body subject to tension test at different constant rates of strain.

If the Kelvin body is subjected to a constant rate of strain  $\dot{\mathbf{E}}_0 = \mathbf{c}$ , the strain  $\mathbf{E}_0 = \mathbf{ct}$ ; hence the stress-strain tensor relation,

$$\mathbf{T}_0 = 2G\mathbf{E}_0 + 2\eta\mathbf{c} \quad (35 \cdot 10)$$

is linear throughout, the tangent modulus being constant and equal to  $2G$ ; for  $\mathbf{E}_0 = 0$  the initial stress tensor  $\mathbf{T}_{00} = 2\eta\mathbf{c}$  (Fig. 35·2). For  $\dot{\mathbf{E}}_0 = \text{const.}$ ,  $\mathbf{E}_0 = 0$ ; for this condition the

Kelvin body is perfectly elastic.

For a constant rate  $s_0$  of stress  $\mathbf{T}_0 = s_0 t$ ; the integration of eq. 35·3 gives the strain-time relation,

$$\mathbf{E}_0 = \frac{s_0 t}{2G} \left[ 1 - \frac{\tau}{t} (1 - e^{-t/\tau}) \right] + \mathbf{E}_{00} e^{-t/\tau} \quad (35 \cdot 11)$$

which, for  $s_0 = 0$ , gives the strain-recovery function (elastic after-effect):

$$\mathbf{E}_0 = \mathbf{E}_{00} e^{-t/\tau} \quad (35 \cdot 12)$$

If  $\mathbf{E}_0 = 0$ , for  $t = 0$  the stress-strain tensor relation becomes

$$\mathbf{E}_0 = \frac{\mathbf{T}_0}{2G} \left[ 1 - \frac{s_0 \tau}{\mathbf{T}_0} (1 - e^{-\mathbf{T}_0/\tau s_0}) \right] \quad (35 \cdot 13)$$

The inverse tangent modulus,

$$\frac{d\mathbf{E}_0}{dT_0} = \frac{1}{2G} (1 - e^{-T_0/\tau s_0}) \quad (35 \cdot 14)$$

for  $T_0 = 0$  becomes  $(d\mathbf{E}_0/dT_0)_{T_0=0} = 0$ .

Hence, all stress-strain diagrams have a common vertical tangent at the origin, which is the axis of ordinates; for  $T_0 \rightarrow \infty$  they become parallel, the inclination of the asymptotes being  $\left(\frac{dT_0}{d\mathbf{E}_0}\right)_{T_0 \rightarrow \infty}$   $= 2G$  (Fig. 35.3). On the ordinate axis these asymptotes cut off the values  $(\tau s_0)$ .

A constant stress tensor  $\mathbf{T}_0 = \mathbf{T}_{00}$  and, consequently,  $\dot{\mathbf{T}}_0 = 0$ , acting on the Maxwell body, produces a constant strain rate and a steadily increasing strain tensor, since

$$\dot{\mathbf{E}}_0 = \frac{1}{2\eta} \mathbf{T}_{00}$$

and

$$\mathbf{E}_0 = \frac{1}{2\eta} \mathbf{T}_{00}t + \text{const} \quad (35 \cdot 15)$$

The constant is the strain  $\mathbf{E}_{00}$  at  $t = 0$ , which is the elastic strain produced by  $\mathbf{T}_{00}$ . Hence,

$$\mathbf{E}_0 = \frac{1}{2G} \mathbf{T}_{00} + \frac{1}{2\eta} \mathbf{T}_{00}t = \frac{1}{2G} \mathbf{T}_{00} \left[ 1 + \frac{t}{\tau} \right] \quad (35 \cdot 16)$$

If a constant strain  $\mathbf{E}_0 = \mathbf{E}_{00}$  is applied and sustained, eq. 35.4 with  $\dot{\mathbf{E}}_0 = 0$  produces the relation:

$$\mathbf{T}_0 = \mathbf{T}_{00}e^{-(G/\eta)t} = 2G\mathbf{E}_{00}e^{-t/\tau} \quad (35 \cdot 17)$$

Hence, if the deformation is kept constant the initially induced stress due to the strain tensor  $\mathbf{E}_{00}$  is gradually relaxed. The rate of stress relaxation is governed by the *relaxation time*  $\tau = \eta/G$ . Since

$$\dot{\mathbf{T}}_0 = -\mathbf{T}_{00} \frac{1}{\tau} e^{-t/\tau} \quad (35 \cdot 18)$$

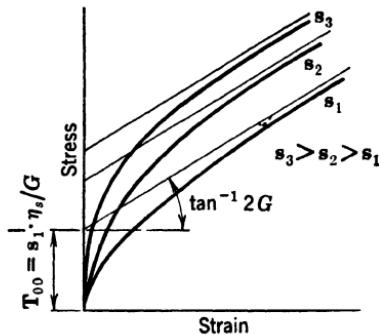


FIG. 35.3 Stress-strain relation for Kelvin body subject to tension test at different constant rates of stress.

the stress rate at  $t = 0$ ,

$$(\dot{\mathbf{T}}_0)_{t=0} = -\frac{1}{\tau} \mathbf{T}_{00} \quad (35 \cdot 19)$$

Hence the tangent at  $t = 0$  to the stress-time diagram of the Maxwell body (Fig. 35·4) cuts off the relaxation time on the time axis.

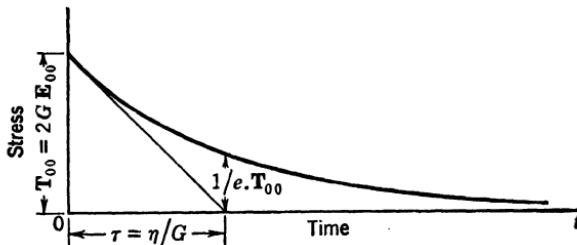


FIG. 35·4 Stress-time curve for Maxwell body under constant strain (relaxation diagram).

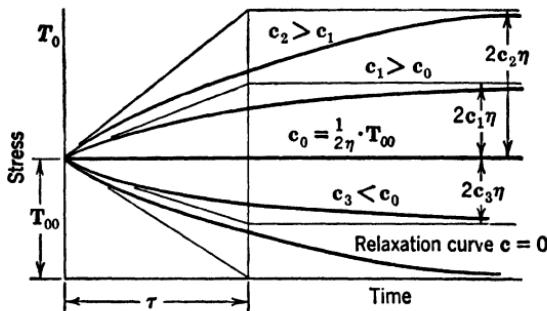


FIG. 35·5 Stress-time diagrams for Maxwell body subject to tension test at different strain rates after having been initially stressed to  $T_{00}$  (after Reiner<sup>35·2</sup>).

If a constant rate of strain  $\dot{\mathbf{E}}_0 = \mathbf{c}$  is applied, the relation

$$\mathbf{T}_0 = \mathbf{T}_{00} e^{-t/\tau} + 2\mathbf{c}\eta(1 - e^{-t/\tau}) \quad (35 \cdot 20)$$

is obtained from eq. 35·4. With  $\mathbf{c} = 0$  eq. 35·15 is transformed into the relaxation function 35·12; for different values of  $\mathbf{c}$  a family of stress-time curves can be drawn (Fig. 35·5). If the strain rate  $\mathbf{c}_0 = \frac{1}{2\eta} \mathbf{T}_{00}$ , eq. 35·15 becomes  $\mathbf{T}_0 = \mathbf{T}_{00} = \text{const}$ ,

and the respective stress-time curve becomes a parallel to the  $t$  axis. For all strain rates  $c > c_0$ , the stresses increase with time, whereas, for  $c < c_0$ , the stresses gradually decrease. The strain rate  $c = c_0$  defines the limiting condition for which the rate of energy dissipation just equals the rate of energy application.<sup>35·2</sup>

The stress-strain relation of the Maxwell body subject to a constant strain rate is obtained by introducing  $\dot{\mathbf{E}}_0 = c$  or  $\mathbf{E}_0 = ct$ , into eq. 35·15, assuming that for  $\mathbf{E}_0 = 0$ ,  $\mathbf{T}_0 = \mathbf{T}_{00} = 0$ .

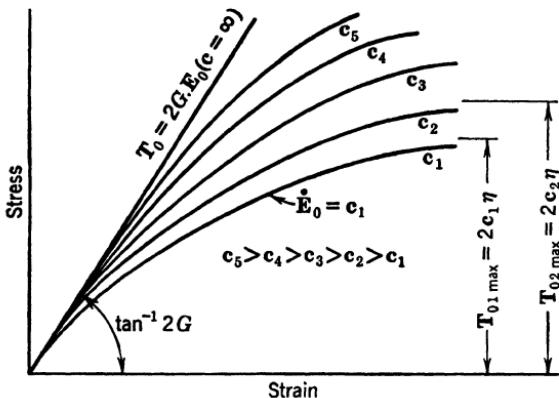


FIG. 35·6. Stress-strain relation of Maxwell body subject to tension test at different constant strain rates  $\dot{\mathbf{E}}_0 = c$ .

The elimination of  $t$  results in the stress-strain function,

$$\mathbf{T}_0 = 2c\eta(1 - e^{-\mathbf{E}_0/c\tau}) \quad (35\cdot21)$$

the tangent modulus,

$$\frac{d\mathbf{T}_0}{d\mathbf{E}_0} = 2Ge^{-\mathbf{E}_0/c\tau} \quad (35\cdot22)$$

for  $\mathbf{E}_0 = 0$  becomes  $(d\mathbf{T}_0/d\mathbf{E}_0)_{(\mathbf{E}_0=0)} = 2G$ .

Thus, the tangent at the origin of the family of stress-strain functions is the stress-strain relation of the linear elastic body; this relation also holds for the whole range of stress if  $c = \infty$ . For different strain rates the deviation of the respective stress-strain function from linear elastic behavior is the less, the higher the value of the applied strain rate (Fig. 35·6).

If after attaining a certain level of strain  $\mathbf{E}_{00}$  the strain rate is

reversed, the equation of the unloading branch of the stress-strain tensor relation may be written in the form:

$$\mathbf{T}_0 = 2c\eta e^{-\mathbf{E}_{00}/cr} \cdot (e^{\mathbf{E}_0/cr} - 1) \quad (35 \cdot 23)$$

valid for  $\mathbf{E}_{00} > \mathbf{E}_0 > 0$ ; its tangent modulus:

$$\frac{d\mathbf{T}_0}{d\mathbf{E}_0} = 2Ge^{-\mathbf{E}_{00}/cr} \cdot e^{\mathbf{E}_0/cr} \quad (35 \cdot 24)$$

for  $\mathbf{E}_0 = \mathbf{E}_{00}$  becomes  $(d\mathbf{T}_0/d\mathbf{E}_0)_{\mathbf{E}_0 = \mathbf{E}_{00}} = 2G$ .

The tangents at the origin of the loading diagrams are necessarily parallel to the tangents at  $\mathbf{E}_0 = \mathbf{E}_{00}$  of the unloading diagrams (Fig. 35·7).

If the Maxwell body is subject to a constant rate  $s_0$  of stress  $\mathbf{T}_0 = s_0 t$ ; the stress-strain relation:

$$\begin{aligned} \mathbf{E}_0 &= \frac{1}{2G} s_0 t + \frac{1}{2} s_0 \frac{t^2}{2\eta} \\ &= \frac{1}{2G} \mathbf{T}_0 \left( 1 + \frac{\mathbf{T}_0}{2\tau s_0} \right) \end{aligned} \quad (35 \cdot 25)$$

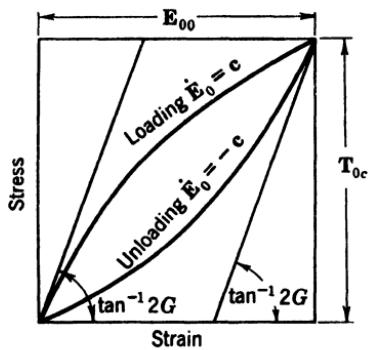


FIG. 35·7 Stress-strain curve for loading of a Maxwell body at a strain rate  $\dot{\mathbf{E}}_0 = c$  up to a strain  $\mathbf{E}_{00}$  and unloading it at a strain rate  $\dot{\mathbf{E}}_0 = -c$ .

for  $\mathbf{T}_0 = 0$  becomes  $(d\mathbf{E}_0/d\mathbf{T}_0)_{(\mathbf{T}_0=0)} = 1/2G$ . Again, the tangent at the origin of the stress-strain curves is the relation of linear elasticity, which holds also for the whole range of strain if  $s_0 = \infty$ . For different stress rates a family of parabolic stress-strain functions is obtained; again, they deviate from linear elasticity the less, the higher the strain rate applied (Fig. 35·8). However, the stresses of the constant-strain-rate stress-strain curves approach limiting values, whereas those of the constant-stress-rate curves increase monotonously.

The foregoing relations connect the deviators of stress and stress rate with the deviators of strain and of strain velocity. The relations connecting stress and strain in simple mechanical

The equation of the inverse tangent modulus,

$$\frac{d\mathbf{E}_0}{d\mathbf{T}_0} = \frac{1}{2G} \left( 1 + \frac{\mathbf{T}_0}{\tau s_0} \right) \quad (35 \cdot 26)$$

tests must be derived by resolving the tensors of stress and strain with regard to the particular conditions of the test.

Since in shear (for instance in the torsion test of very thin-walled tubes) the spherical tensors of both stress and strain vanish, the deviator relation is identical with the relation between the shear stress and strain components. Hence, the deviator relation for the Maxwell body, for instance, can be directly transformed into the shear stress-strain relation,

$$\dot{g}_{12} = \frac{1}{2G} \dot{s}_{12} + \frac{1}{2\eta} s_{12} \quad (35 \cdot 27)$$

as all the constants remain unchanged.

Axial load tests, however, require resolution of the tensors. By transforming the matrix of the deviator of strain so as to make it comparable to the matrix of the deviator of stress,<sup>35.3</sup> the deviator relation of the linear elastic body is transformed into the stress-strain relation in uniaxial loading,

$$s = e_1 2G(1 + \mu) \quad (35 \cdot 28)$$

The equation of the volume-constant Maxwell body in terms of uniaxial stress and strain can thus be directly written in the form,

$$\dot{e}_1 = \frac{1}{3G} \dot{s}_1 + \frac{1}{3\eta} s_1 = \frac{1}{E} \dot{s}_1 + \frac{1}{\lambda} s_1 \quad (35 \cdot 29)$$

where  $\lambda = 3\eta$  is the coefficient of viscosity for uniaxial volume-constant deformation.

**ELASTIC-VISCOELASTIC ANALOGY.** If the deviators of stress and strain are developed into series of their time derivatives, the general stress-strain relations of volume-constant linear viscoelastic materials, may be written in the general form,

$$P\mathbf{T}_{0,jk} = 2Q\mathbf{E}_{0,jk} \quad (35 \cdot 30)$$

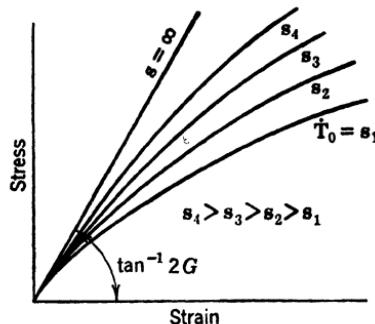


FIG. 35.8 Stress-strain relation for Maxwell body subject to tension test at different constant rates of stress  $\mathbf{T}_0 = \mathbf{s}$ .

where  $P$  and  $Q$  denote, respectively, the linear operators  $a_k \cdot \frac{\partial^k}{\partial t^k}$

and  $b_m \cdot \frac{\partial^m}{\partial t^m}$ , and  $a_k$  and  $b_m$  are constants characteristic of the material.

Alfrey has shown that an analogy exists between elastic and linear viscoelastic materials, which simplifies the solution of boundary problems of volume-constant viscoelastic materials.<sup>35-4</sup> These boundary problems may be of two types:

1. Surface forces are given as functions of space and time; the pertaining stress distribution within the body and the deformations shall be determined.

2. Displacements of the surface are given as functions of space and time; the pertaining displacement and stress distribution within the body shall be determined.

In the solution of the problem, it is assumed that, at the time  $t = 0$ , either the given surface forces and the stress components which are to be determined vanish together with their  $(k - 1)$  time derivatives, or the given surface displacements, as well as the displacement components in the interior which are to be determined vanish together with their  $(m - 1)$  time derivatives.

The elastic-viscoelastic analogy shows that the distribution of deviatoric (distortional) stresses within any volume-constant linear viscoelastic material described by a general relation of the form 35.30 is identical with that of a volume-constant elastic material subject to the same instantaneous surface forces. Thus, if the given surface forces can be written in the form,

$$f_i = f_i(x)p(t) \quad (35.31)$$

where  $x$  denotes the space coordinates, the stress distribution produced within the linear viscoelastic material

$$s_{ik}(x, t) = s_{0ik}(x)p(t) \quad (35.32)$$

where  $s_{0ik}(x)$  denotes the deviatoric stresses produced by the surface forces  $f_i(x)$  within a volume-constant linear elastic body of identical shape. According to eq. 35.30, the strain components  $e_{ik}$  in the viscoelastic body can be written in the form,

$$e_{ik}(x, t) = e_{0ik}(x)r(t) \quad (35.33)$$

where  $r(t)$  the response function satisfies the differential equation,

$$Qr(t) = Pp(t) \quad (35 \cdot 34)$$

and vanishes with its derivatives up to the  $(m - 1)$  order for  $t = 0$ . The quantity  $e_{0ik}(x)$  is connected with  $s_{0ik}(x)$  by the relation,

$$s_{0ik} = 2e_{0ik} \quad (35 \cdot 35)$$

The *equivalent* elastic strain  $e_{0ik}$  is thus the deviatoric strain in an elastic body of unit modulus of rigidity, subjected to the surface forces  $f_i(x)$ .

The response function  $r(t)$  of any viscoelastic material is easily obtained from the consideration of the effect of a simple shearing stress  $s = 2p(t)$ . The shearing strain produced by this stress and obtained by solving the basic equation of the material for pure shear is  $r(t)$ . The strain within the viscoelastic body produced by the surface forces  $f_i(x)p(t)$  can now be computed by multiplying the equivalent elastic strain  $e_{0ik}$  by the response function  $r(t)$ . Solutions of this type may be superimposed because the differential equations for stresses and strains are linear.

A similar procedure is adopted for the determination of stresses in the second boundary value problem of viscoelastic materials. If the surface displacements are given in the form,

$$u_i = u_i(x)p(t) \quad (35 \cdot 36)$$

and  $s_{0ik}$  denotes the *equivalent* elastic deviatoric stress produced by these displacements in the linear elastic body, the stress distribution in the viscoelastic body is given by

$$s_{ik}(x, t) = s_{0ik}(x)r(t) \quad (35 \cdot 37)$$

where  $r(t)$  is the response function to be obtained as before.

Because of the elastic-viscoelastic analogy it is possible to use linear viscoelastic materials in the photoelastic investigations of the elastic distribution of stresses. However, the strain-optical relations vary with the type of viscoelastic behavior and are different for a material with after-effect and for a material with both after-effect and creep, since the birefringence arises solely from the deformation of the elastic phases of the material. These relations have been derived by Mindlin.<sup>35·5</sup>

### 36. Mechanical Models. Simple Behavior

In order to account for the inelastic effects in materials in terms of their structure it is necessary to analyze mechanisms to which these effects may be due. A promising line of attack of this problem would be the setting up of *structural* models which could explain in simple terms the structure and its mechanism of deformation. However, the development of such models would require a knowledge of the internal structure of the material, which is frequently lacking. In order to overcome this difficulty and to obtain from the model qualitative information concerning the internal structure of the considered material, *mechanical* models have been developed. The principal value of such mechanical models is that they show how the different types of simple behavior may be superimposed; they provide thus the clues to the development of real *structural* models.

In the study of molecular or atomic mechanisms of inelastic deformation the analysis of a mechanical model will frequently be more effective than that of a stress-strain relation. In a model analysis each contribution to the stress or strain of any one element may actually be identified with some specific molecular or atomic process so that the individual strain or stress component may be assumed to possess a certain physical significance.

The theory of mechanical models considers combinations of elements which are supposed to behave mechanically like the constituent phases of the material, but which, apart from this mechanical behavior, have nothing in common with the real material. The mechanical models devised to illustrate inelastic behavior are usually built up of combinations of two model-elements representing the two basic types of deformation (Fig. 36·1):

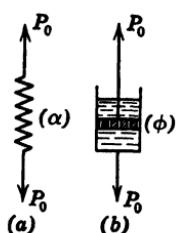


FIG. 36·1 Simple model elements: (a) spring (elastic) and (b) dashpot (viscous).

2. A *dashpot*, consisting of a perforated piston moving in a cylinder containing a viscous liquid, for viscous deformation.

The spring element, which obeys Hooke's law, is a model of the linear elastic body; if the liquid in the cylinder obeys Newton's

1. A perfectly elastic spring for elastic deformation.

law of viscosity, the dashpot is a model of the Newtonian liquid. These elements can be combined in different ways so as to approximate the observed mechanical behavior of real materials. However, as long as models are limited to combinations of the two elements of linear elasticity and of linear viscosity, they can only reproduce the deformation of linear viscoelastic materials.

Mechanical models can only be correlated with the behavior of materials under simple conditions of testing, such as pure shear or volume-constant homogeneous stress or strain.

For the spring element the relation between displacement  $x$  and force  $P$  is

$$x_1 = \alpha P \quad (36 \cdot 1)$$

where  $\alpha$  is the spring constant. For the dashpot the relation between the rate of displacement of the piston  $dx_2/dt$  under the action of a force  $P$  is

$$\frac{dx_2}{dt} = \phi P \quad \text{and} \quad x_2 = \phi \int_0^t P dt \quad (36 \cdot 2)$$

where  $\phi$  is the reciprocal value of the viscosity coefficient, called the *fluidity*.

There are two possibilities of combining elements; they can be coupled in either series or parallel. If the two elements are coupled in series (Fig. 36·2), the force acting in them will be the same and equal to the total applied force. Hence, the deformation is the total change of length produced in the individual elements,

$$x = x_1 + x_2 = \alpha P + \phi \int_0^t P dt \quad (36 \cdot 3)$$

or, by differentiation,

$$\frac{dx}{dt} = \phi P + \alpha \frac{dP}{dt} \quad (36 \cdot 4)$$

A comparison of eq. 36·4 with that of the Maxwell body (Table 35·1) shows that by coupling the basic elements in series a model of the Maxwell body is produced. The relaxation time

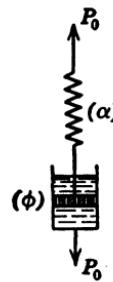


FIG. 36·2 Simple elements combined in series (Maxwell unit).

of the system is  $\tau = \alpha/\phi$ , and

$$P = e^{-t/\tau} \cdot \left( P_0 + \frac{1}{\alpha} \int_0^t \frac{dx}{dt} e^{t/\tau} dt \right) \quad (36 \cdot 5)$$

where  $P_0 = (P)_{t=0}$ .

If the two elements are coupled in parallel (Fig. 36·3), their extensions under load must be the same. The force acting on the spring is given by

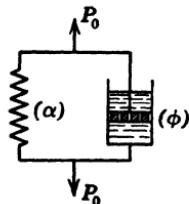


FIG. 36·3 Simple elements combined in parallel (Kelvin unit).

$$P_1 = \frac{x}{\alpha} \quad (36 \cdot 6)$$

and the force on the dashpot by

$$P_2 = \frac{1}{\phi} \cdot \frac{dx}{dt} \quad (36 \cdot 7)$$

If both equations are combined, the load  $P$  on the system,

$$P = P_1 + P_2 = \frac{x}{\alpha} + \frac{1}{\phi} \frac{dx}{dt} \quad (36 \cdot 8)$$

When eq. 36·8 is compared with that of the Kelvin body, it is obvious that by coupling of two basic elements in parallel a model of the Kelvin body is produced. This is also the model of Kelvin's "sponge, filled by a viscous liquid";<sup>35·5</sup> its behavior can therefore be expected to approximate the mechanical behavior of two-phase materials, consisting of a "springy" continuous structure, the voids of which are filled by a viscous liquid. The integral of the model equation is

$$x = e^{-t/\tau} \cdot \left( x_0 + \phi \int_0^t P e^{t/\tau} dt \right) \quad (36 \cdot 9)$$

where  $x_0 = (x)_{t=0}$ .

In the case of a load of periodically varying magnitude the viscous resistance causes a phase difference between applied force and resulting deformation, associated with dissipation of mechanical work into heat. The load deformation diagram of the Maxwell body is a closed ellipse the area of which  $W = \pi P_0^2 \phi / \omega$  represents the energy dissipated per load cycle;  $P_0$  denotes the load amplitude and  $\omega$  its frequency. For the Kelvin body the

stable closed elliptical load-deformation diagram is only reached after a few cycles of transient behavior.

In discussing mechanical models it has been assumed so far that inelastic deformation is produced by any load intensity. If the behavior of plastic materials is to be analyzed in terms of mechanical models, the yield limit must be introduced as an additional constant. This can be done either

1. By introducing a friction resistance between piston and cylinder of the dashpots, which would prevent the pistons from moving as long as the force on the dashpot remains below the yield limit.

2. By introducing a third model element consisting of a weight resting on a plane connected to a spring, with a friction force equal to the yield limit, preventing the movement of the block under the action of forces below this limit (Fig. 36·4).

The new element can be considered a St. Venant unit, since the weight remains at rest and the deformation is elastic until the force exceeds the yield limit, at which a time-independent motion along the plane is initiated. The Maxwell unit with friction between piston and cylinder of the dashpot, on the other hand, reproduces the behavior of a Bingham body with elasticity. By constructing a composite model, consisting of a number of such units with different friction values in the dashpots, coupled in parallel, the plasticity of metals can be reproduced.<sup>36·1</sup>

### 37. Mechanical Models. Complex Behavior

Real materials are usually of more complex behavior than that which can be described by simple linear models. Such behavior is frequently approximated by the coupling of more than two basic elements, or by the coupling of several Maxwell units or Kelvin units.

By combining, for instance, a Maxwell unit and a Kelvin unit in series and subjecting it to a constant load  $P_0$  acting between  $t = 0$  and  $t = t_0$ , the total elongation is obtained by adding that of the two units the constants of which are marked by the subscripts  $m$  (Maxwell) and  $k$  (Kelvin), respectively.

$$x = P_0[\alpha_m + \phi_m t + \alpha_k(1 - e^{-t/\tau_k})] + x_0 e^{-t/\tau_k} \quad (37\cdot1)$$

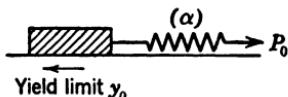


FIG. 36·4 St. Venant model unit.

where  $x_0$  denotes the deformation imposed at  $t = 0$ . This equation shows that the total deformation under load is made up of three parts: instantaneous elastic deformation, linear creep, and after-effect. If at  $t = t_0$  the load is released, of the total deformation produced by the constant force  $P_0$ ,

$$\begin{aligned}x_{(t=t_0)} &= P_0\alpha_m + P_0\phi_m t + P_0\alpha_k(1 - e^{-t_0/\tau_k}) \\&= x_{\text{el.}} + x_{\text{perm.}} + x_{\text{after-effect}}\end{aligned}\quad (37 \cdot 2)$$

the elastic part will be recovered immediately, the anelastic part will be recovered gradually, and the creep will remain permanently (Fig. 37·1).

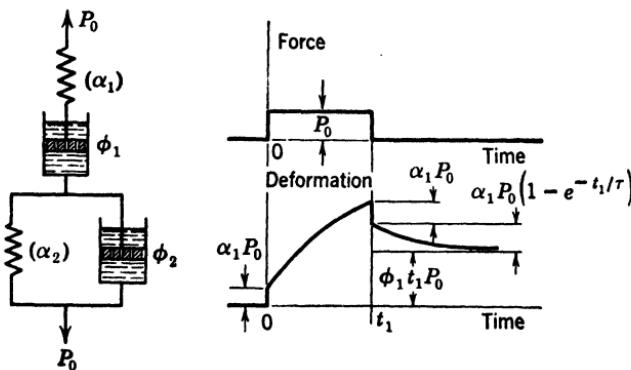


FIG. 37·1 Load-time and elongation-time diagram of model consisting of Maxwell unit and Kelvin unit coupled in series.

If a condition is now considered to exist such that in the course of unloading, at a time  $t_1 > t_0$ , the system is suddenly loaded by a force  $-P_1$  of very short duration,<sup>37·1</sup> acting in the direction opposite to that of the action of  $P_0$ , this force will act mainly on the spring of the Maxwell unit and cause a decrease of  $x$  by  $\alpha_m P_1$ . This deformation is almost fully recovered by removal of  $-P_1$ . From that moment on, the original after-effect proceeds as before, but with a slightly lower horizontal asymptote, because of the slight permanent deformation in the opposite direction produced by  $-P_1$  in the dashpot of the Maxwell unit (Fig. 37·2). The system has thus developed a "memory"; one part of it responds to processes of short duration by contracting and extending in response to the load  $P_1$ ; the other part acts on

the "memory" of the previous load by proceeding to contract, at least for a certain time, in accordance with the after-effect produced by  $P_0$ .

"Memory" effects of this type may be produced experimentally by twisting wires, filaments, or threads of suitable materials beyond the limit of elastic deformation. If a torque is applied in one direction, the after-effect tends to reverse the resulting twist after the torque has been released. If during the period of recovery of the twist a second torque is applied in the opposite direction and released, a second after-effect is produced. Both after-effects are superimposed, and the rates and directions of the resulting after-effect which determine the visible behavior of the wire will depend on the respective times of application and values of the different torques (Fig. 37·3).

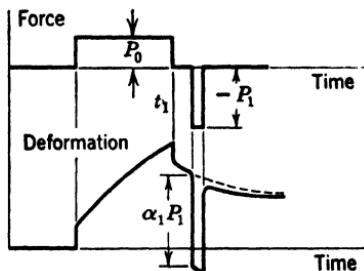


FIG. 37·2 Memory effect produced in model represented in Fig. 37·1 (after Burgers<sup>37·1</sup>).

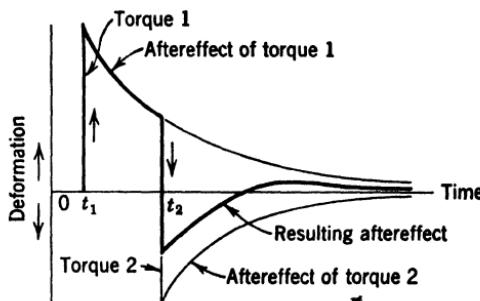


FIG. 37·3 Superimposed after-effects of torques of opposite direction.

If a number of torques of varying intensity and duration are applied consecutively in both directions, the resulting after-effects may produce very erratic observable deformations, all relations between applied torque and resulting twist and after-effect having become intraceable because of the superposition of the various after-effects. "Memory" tests on rubber and glass

threads and on metal wires have been performed by various investigators.<sup>37-2</sup>

It has been found that the behavior of real materials usually cannot be described by two constants, the elastic modulus and the relaxation time alone. The introduction of a number of constants, however, makes it possible to reproduce rather complex behavior. This is done either by parallel coupling of a number  $n$  of Maxwell units or by a coupling in series of a number  $n$  of Kelvin units (Fig. 37·4). Each unit is defined by its con-

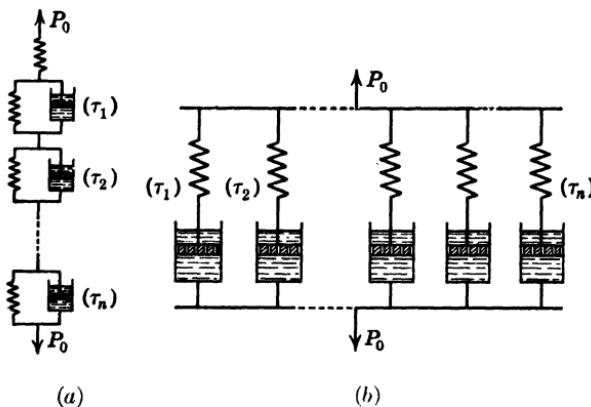


FIG. 37·4 Coupling of (a) Maxwell elements and (b) Kelvin elements for representation of complex inelastic behavior.

stants  $\alpha_{mi}$ ,  $\phi_{mi}$  or  $\alpha_{ki}$ ,  $\phi_{ki}$ , the subscript  $i$  taking, consecutively, all values from 1 to  $n$ . Both types of superposition of units have been applied in the analysis of textile fibers, glass, and high polymers.<sup>37-3</sup>

The coupling in parallel of a number of Maxwell units is equivalent to the structural assumption of the coexistence of a number of phases with different relaxation times  $\tau_{mi} = \alpha_{mi}/\phi_{mi}$ . This assumption has a certain physical justification since in materials of unordered structure the existence of phases of different relaxation times may be expected as a result of:

1. The difference in the size of particles, especially in molecular compounds.
2. Different relaxation rates in an otherwise homogeneous assembly of particles of the same size, due to variations in the activation energy of particles in their respective equilibrium positions.

If a system consists of two Maxwell units coupled in parallel with the constants  $\alpha_1$ ,  $\alpha_2$ ,  $\phi_1$ ,  $\phi_2$ , and the respective relaxation times  $\tau_1 = \alpha_1/\phi_1$  and  $\tau_2 = \alpha_2/\phi_2$ , its response to a constant load applied at time  $t = 0$  is determined by the conditions, that (1) the sum of the forces  $P_1$  and  $P_2$  in the two elements is equal to the applied force  $P_0$ , and (2) the deformation of both Maxwell units under their respective loads  $P_1$  and  $P_2$  must be equal. These conditions produce the relation,<sup>37·4</sup>

$$x = P_0 p [1 + q(1 - e^{-t/\tau}) + r \cdot t] \quad (37 \cdot 3)$$

where

$$p = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)}$$

$$q = - \left(1 - \frac{\tau}{\tau_1}\right) \left(1 - \frac{\tau}{\tau_2}\right)$$

$$r = \frac{\tau}{\tau_1 \tau_2}$$

and

$$\tau = \frac{\alpha_1 + \alpha_2}{\phi_1 + \phi_2}$$

Equation 37·3 has the same general form as eq. 37·2, containing an elastic component, a uniform creep component, and an after-effect component. The instantaneous deformation is given by  $P_0 p$ ; this value, multiplied by  $(rt)$ , gives the uniform creep component; the term  $P_0 p q (1 - e^{-t/\tau})$  represents the deformation due to redistribution of the load as a result of the coexistence of two relaxation times.

After removal of the load at time  $t_1$  the residual deformation at a subsequent time  $t$  is given by the expression,

$$x = P_0 p [rt_1 + q \cdot e^{-t/\tau} (e^{t_1/\tau} - 1)] \quad \text{for } t > t_1 \quad (37 \cdot 4)$$

it consists of a permanent set  $P_0 p r t_1$  and a transient residual deformation (after-effect).

The load-time and deformation-time diagrams drawn in Fig. 37·5 show that under constant load the material flows at a gradually decreasing rate, a phenomenon that is usually described as transient or "primary" creep, until it reaches the steady state of linear viscous or "secondary" creep. It may therefore be con-

cluded that the introduction of different relaxation times by the coupling of Maxwell units is sufficient to produce transient creep.

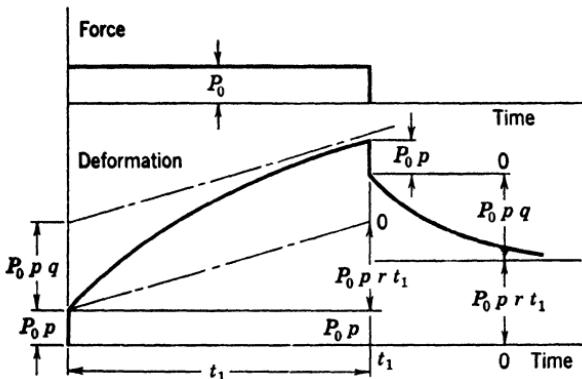


FIG. 37-5 Force-time and deformation-time diagram of model consisting of two Maxwell units coupled in parallel.

If the material, the behavior of which is to be reproduced by the model, consist of a continuous solid skeleton, at least one

elastic element must be coupled in parallel with the Maxwell elements. The external load will then, through the relaxation of the Maxwell units, be gradually transferred into the elastic element. After load release there will, however, be no permanent deformation, only after-effect. Such behavior is obtained if in the model consisting of two Maxwell units one relaxation time approaches infinity (Fig. 37-6).

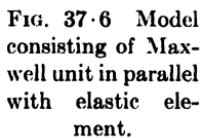


FIG. 37-6 Model consisting of Maxwell unit in parallel with elastic element.

Under a constant load  $P_0$  applied at time  $t = 0$  the deformation of such a model at time  $t$ ,

$$x = P_0 \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2} \left[ 1 + \frac{\alpha_1}{\alpha_2} (1 - e^{-t/\tau}) \right] \quad (37-5)$$

where  $\tau = \frac{\alpha_1 + \alpha_2}{\phi_2}$ . If the load is removed at  $t = t_1$ , the residual deformation at a subsequent time  $t$ ,

$$x = P_0 \frac{\alpha_1^2}{\alpha_1 + \alpha_2} [(1 - e^{-t/\tau}) - (1 - e^{-(t-t_1)/\tau})] \quad (37-6)$$

The force-time and deformation-time diagrams are shown in Fig. 37·7. After the occurrence of the instantaneous elastic deformation in both springs, the load carried by the Maxwell unit is gradually transferred to the spring. Upon load release the spring cannot recover its initial length instantaneously because of the resistance of the dashpot in the Maxwell unit. However, the remaining tension in the spring is gradually relaxed as a result of the viscous flow in the dashpot, and the initial length of the model is asymptotically recovered.

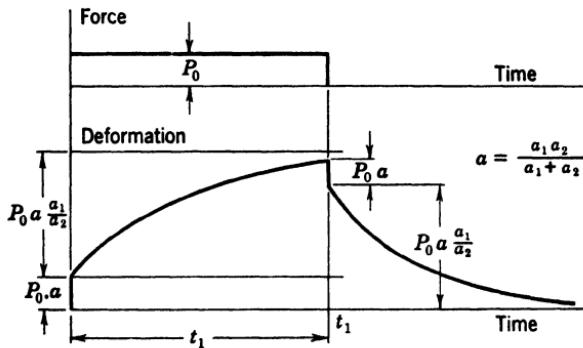


FIG. 37·7 Force-time and force-deformation diagram of model represented in Fig. 37·6.

If a periodic force  $P = P_0 \sin \omega t$  is imposed, the behavior of this model under two limiting conditions of frequency can be predicted without mathematical analysis. Periodic forces of very high frequency will produce simultaneous elastic deformation of both springs in phase with the force, the dashpot of the Maxwell unit remaining practically rigid. Thus,

$$x = \alpha P_0 \sin \omega t \quad (37\cdot7)$$

where  $\alpha = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ . If, on the other hand, the periodic force

is applied very slowly, the Maxwell unit will always be very nearly relaxed, the force being carried by the spring element alone. The deformation is therefore again in phase with the applied force, but the amplitude increases to

$$x = \alpha_1 P_0 \sin \omega t \quad (37\cdot8)$$

Hence, as the frequency of the applied force decreases, the

apparent rigidity of the system decreases steadily. However, the angle of lag between force and deformation, which is zero under both limiting conditions, must necessarily first increase and then decrease, attaining a maximum between the limiting conditions defined by eqs. 37·7 and 37·8.

Models with multiple relaxation times can be made to represent the deformation of amorphous materials, such as glass, rather closely. Various functions describing "primary" creep or after-effect can be produced by parallel coupling of Maxwell units. The larger the number of units coupled, the more extended the period of "primary" creep and the better the approximation to given creep curves that can be accomplished.

If a model of a number  $n$  of relaxing (Maxwell) elements and one spring element is subject to an instantaneously applied deformation  $x_0$ , the initial load in the system,

$$P_0 = x_0 \sum_1^n \frac{1}{\alpha_i} \quad (37 \cdot 9)$$

the load in an individual unit

$$P_{i0} = \frac{x_0}{\alpha_i} \quad (37 \cdot 10)$$

Since after a time  $t$  this load has decreased to

$$P_i = P_{i0} e^{-t/\tau_i} \quad (37 \cdot 11)$$

the total load in the system at time  $t$  has also decreased to

$$P = \frac{x_0}{\alpha_1} + x_0 \sum_2^n \frac{(e^{-t/\tau_i})}{\alpha_i} \quad (37 \cdot 12)$$

where  $\alpha_1$  denotes the constant of the spring element.

The analysis of the behavior of a multiple-element model is rather cumbersome, even for constant load, since a complex redistribution of loads among the Maxwell units takes place. Immediately on load application the Maxwell units with very short relaxation times begin to give up their loads. Since the units with long relaxation times give up their loads very slowly, they will temporarily carry some of the load given up by the other units. Hence, their loads will first increase and then

decrease, whereas for units with short relaxation times the load decreases continuously to zero; in the spring elements it increases continuously to its final value associated with the full load.

### 38. Superposition Theories<sup>38.1</sup>

If it can be assumed that no permanent structural changes occur in a material during deformation under constant load, the deformation at any instant can be divided into two parts: the instantaneous deformation, which is proportional to the stress acting at that instant, and the delayed deformation, which is a function of the loading history. The total deformation, as a function of time can thus be written in terms of a model:

$$x(t) = \alpha P_0 [1 + \beta_c \psi_c(t)] \quad (38 \cdot 1)$$

where  $\beta_c$  is a constant of proportionality and  $\psi_c(t)$  a function of time, which increases from zero at time  $t = 0$  to a finite value at infinite time. The function  $\psi_c(t)$  for deformation under constant load is called the *creep function* and represents the delayed component of the deformation.

On the other hand, if a constant deformation  $x_0$  is applied at zero time, the force required to sustain this deformation can be divided into two parts: one, representing the initial force required to produce instantaneously the deformation  $x_0$ , the other representing the proportionate decay of the force with time under constant deformation. The force, as a function of time, can thus be written in the form,

$$P(t) = \frac{x_0}{\alpha} [1 - \beta_r \psi_r(t)] \quad (38 \cdot 2)$$

where  $\beta_r$  is a constant of proportionality and  $\psi_r(t)$  a function of time, which increases from zero at time  $t = 0$  to a finite value smaller than  $1/\beta_r$  at infinite time; this function is called the *relaxation function*.

For a Maxwell unit the relaxation function is obtained from the comparison of eqs. 38·2 and 36·5 for  $P_0 = x_0/\alpha$  and  $dx/dt = 0$ ,

$$\frac{x_0}{\alpha} [1 - \beta_r \psi_r(t)] = \frac{x_0}{\alpha} e^{-t/\tau} = \frac{x_0}{\alpha} [1 - (1 - e^{-t/\tau})] \quad (38 \cdot 3)$$

in the form,

$$\beta_r \psi_r(t) = 1 - e^{-t/\tau} \quad (38 \cdot 4)$$

with  $\beta_r = 1$ , whereas the creep function follows from the comparison of eqs. 38·1 and 36·3 for  $P = P_0$ :

$$\beta_c \psi_c(t) = \frac{\phi}{\alpha} t = \frac{t}{\tau} \quad (38 \cdot 5)$$

Since the function  $e^{-t/\tau}$  can be approximated by  $(1 - t/\tau)$  for small values of  $t/\tau$ , the relations  $x(t)/\alpha P$  and  $\alpha P(t)/x_0$ , as defined by the creep and relaxation equations 38·1 and 38·2, respectively, are the mirror images of each other for very small values of  $t/\tau$  and for  $\beta_r = \beta_c = 1$ .

Considering a more complex mechanical model, such as that consisting of a spring and a Maxwell unit coupled in parallel, relaxation and creep functions are again obtained by comparison of the eqs. 38·2 and 38·1 with eq. 37·5 describing the model behavior:

$$\beta_r \psi_r(t) = \frac{\alpha_1}{\alpha_1 + \alpha_2} (1 - e^{-t/\tau}) \quad (38 \cdot 6)$$

with  $\beta_r = \frac{\alpha_1}{\alpha_1 + \alpha_2}$ , and

$$\beta_c \psi_c(t) = \frac{\alpha_1}{\alpha_2} (1 - e^{-t/\tau}) \quad (38 \cdot 7)$$

with  $\beta_c = \alpha_1/\alpha_2$ . In this case the relations  $x(t)/\alpha P_0$  and  $\alpha P(t)/x_0$ , as defined by eqs. 38·1 and 38·2, are two curves that can become the mirror image of each other if and only if  $\beta_r = \beta_c$ . This requires that  $\alpha_1$  be small in comparison with  $\alpha_2$ , which can only be the case if the delayed component of the deformation is small in comparison with the instantaneous component. If both components are of the same order of magnitude, the simple relation between creep and relaxation no longer exists; in such cases the results of creep and of relaxation tests are not directly convertible.

For a model with a large number of relaxation times the comparison of eqs. 37·12 and 38·2 produces the function,

$$1 - \psi_r = \frac{\sum_{i=1}^n \frac{1}{\alpha_i} e^{-t/\tau_i}}{\sum_{i=1}^n \frac{1}{\alpha_i}} \quad (38 \cdot 8)$$

which indicates the decay of stress with time, the constants  $1/\alpha_i$  representing the "weights" of the different relaxation times  $\tau_i$  present in the material or in its model. If the *decay function* 38·8 is plotted against time in a logarithmic scale, curves of sigmoidal (ogive) type, with some irregular features, are obtained (Fig. 38·1). The stress decay is the more continuous, the more relaxation times are present. It is on account of the assumption of *discrete* relaxation times and weights that *several* points of

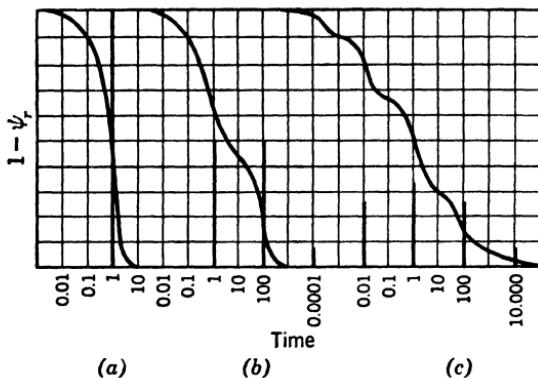


FIG. 38·1 Stress-decay functions ( $1 - \psi_r$ ) for various distributions of relaxation times, plotted against logarithm of time. The heavy vertical lines indicate frequency distribution of relaxation times (after Wiechert<sup>38·2</sup>).

inflection, as shown in Fig. 38·1, are introduced into the decay functions.<sup>38·2</sup> Since in real materials one inflection point only is usually observed, the discontinuous model, consisting of a finite number of Maxwell units, should be transformed into a continuous one, characterized by the continuous distribution function of relaxation times  $F(\tau)$ . The area under the distribution curve must be unity (normalizing condition). If the discrete weights  $1/\alpha_i$  in eq. 38·8 are replaced by the continuous distribution curve, the decay function,

$$1 - \psi_r = \int_0^{\infty} F(\tau) d\tau e^{-t/\tau} \quad (38·9)$$

is obtained. Hence, the relaxation function, in terms of an arbitrary distribution of relaxation times,

$$\psi_r = \int_0^{\infty} F(\tau) (1 - e^{-t/\tau}) d\tau \quad (38·10)$$

The direct computation of the distribution function of relaxation times from observed relaxation curves is very difficult; a more expedient way of analysis consists in writing down an analytical expression for the distribution function  $F(\tau)$ , which appears plausible on theoretical grounds and allows the calculation of the relaxation function, and then comparing and fitting the calculated and the observed curves by changing the parameters of the assumed distribution function of relaxation times.

### 39. Nonlinear Mechanical Models

The derivation of eq. 19·17 from eq. 19·16 shows that linear viscosity represents only a special case of general inelastic behavior. Mechanical models made up of linear dissipative elements therefore cannot be expected to reproduce closely the behavior of real materials of general behavior. Even the combination of linear elements by applying the superposition principle (see Art. 38) is frequently ineffective in reproducing mechanical behavior which is basically nonlinear. This is particularly true for the behavior of polymers at high stress levels as well as for metals.

A mechanical model by the aid of which general inelastic behavior can be reproduced more adequately than by the use of complex linear models is the three-element model shown in Fig. 37·6 in which the linear dissipative element has been replaced by a nonlinear element, the behavior of which is governed by the hyperbolic sine law of general inelasticity (eq. 19·16). This model has been extensively used in textile research.<sup>39·1</sup>

The principal difficulty in the practical use of a nonlinear model is due to the complexity of the mathematical expressions obtained as solutions of the model equation. Even the analysis of a two-element Maxwell unit consisting only of a spring and of a non-linear dashpot leads to expressions that are cumbersome and difficult to use in the actual evaluation of experimental results.

Thus, the equation of the nonlinear Maxwell body for pure shear is obtained by adding the equation of the Hencky solid to that of the general inelastic solid eq. 19·16:

$$\dot{\epsilon} = \frac{1}{2G} \dot{s} + C \sinh(cs) \quad (39\cdot1)$$

or, in terms of model analysis,

$$\dot{x} = \alpha \dot{P} + C \sinh (\beta P) \quad (39 \cdot 2)$$

where  $C$  and  $\beta$  are constants of nonlinear behavior. Under constant load  $P_0$  the nonlinear Maxwell unit deforms at a uniform rate  $\dot{x} = C \sinh \beta P_0$ . If a constant strain rate  $\dot{x} = c_0$  is applied, eq. 39·2 may be written in the form,

$$\frac{d\phi}{d\tau} + \sinh \phi = K \quad (39 \cdot 3)$$

where  $K = c_0/C$ ,  $\phi = \beta P$  and  $\tau = \frac{1}{\alpha} \beta C t$ . For constant strain  $\dot{x} = c_0 = 0$ , and thus  $K = 0$ . Hence, the equation of relaxation of the nonlinear Maxwell unit is

$$\frac{d\phi}{d\tau} + \sinh \phi = 0 \quad (39 \cdot 4)$$

the solution of which for the boundary condition  $\phi = \phi_0$  for  $\tau = 0$ ,

$$\tanh (\phi/2) = \tanh (\phi_0/2) e^{-\tau} \quad (39 \cdot 5)$$

replaces the relaxation equation of the linear Maxwell unit.

The general nonhomogeneous eq. 39·3 which, with different constants, is also the model equation of the three-element nonlinear model represented in Fig. 37·6, has recently been integrated for several loading conditions.<sup>39·1</sup> It has been found that the non-linear dashpot yields rather sharply at a certain intensity of the applied force, independent of the rate of deformation, and does hardly move below this force. Thus the rather abrupt transition between small and large deformations characteristic of metals can be related to the nonlinearity of the dissipative element, making it unnecessary to use a special St. Venant element for the model reproduction of a yield point.

### References

- 33·1 J. H. POYNTING and J. J. THOMSON, *Properties of Matter*, Charles Griffin & Co., London (1934) 134.
- 33·2 E. STERNBERG, *J. Applied Mechanics* 13 (1946) A-53.
- 33·3 M. A. BIOT, *Proc. 5th Intern. Congr. Applied Mech.*, Cambridge, John Wiley & Sons, New York (1939) 117.
- 34·1 R. HOUWINK, *Elasticity, Plasticity, and Structure of Matter*, Cambridge Univ. Press (1937) 69.

34·2 *Ibid.*, 212.  
34·3 H. HENCKY, *J. Rheol.* **2** (1931) 169; *Trans ASME* **55** (1933) 55.  
34·4 P. W. BRIDGMAN, *Proc. Am. Acad. Arts Sci.* **70** (1935) 285; **72** (1938) 207; **63** (1929) 401.  
F. BIRCH, *J. Applied Phys.* **9** (1938) 279.  
34·5 M. MOONEY, *J. Applied Phys.* **11** (1940) 582.  
34·6 G. HAMEL, *Elementare Mechanik*, Teubner, Leipzig (1912) 562.  
34·7 F. D. MURNAGHAN, *Am. J. Math.* **64** (1945) 35.  
B. R. SETH, *Phil. Trans. Roy. Soc. A* **234** (1935) 231.  
34·8 M. REINER, *Am. J. Math.* **70** (1948) 433.  
34·9 D. PANOV, *Compt. rend. acad. sci. URSS* **22** (1939) 158.  
34·10 LORD KELVIN, *Math. & Phys. Papers* **3** (1890) 34.  
35·1 M. REINER, *Ten Lectures on Theoretical Rheology*, R. Mass, Jerusalem (1943) 50.  
35·2 *Ibid.*, 91.  
35·3 *Ibid.*, 46.  
35·4 T. ALFREY, *Mechanical Behavior of High Polymers*, Interscience Pub. New York (1948) 557.  
35·5 R. D. Mindlin, *J. Applied Phys.* **20** (1949) 206.  
36·1 F. C. JENKIN, *Engineer*, **134** (1922) 612.  
37·1 *1st Rept. on Viscosity & Plasticity*, *Roy. Acad. Sci. Amsterdam* (1935) 29.  
37·2 R. KOHLRAUSCH, *Pogg. Ann. Phys.* (3) **12** (1847) 393.  
F. KOHLRAUSCH, *Pogg. Ann. Phys.* (6) **8** (1876) 337.  
F. REHKUH, *Wied. Ann. Phys.* **35** (1888) 476.  
37·3 T. ALFREY, *op. cit.*, 127, 166.  
37·4 H. LEDERMAN, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers*, Textile Foundation, Washington, D. C. (1943) 55.  
38·1 *Ibid.*, 20.  
38·2 E. WIECHERT, *Ann. Phys.* **50** (1893) 335, 546.  
39·1 H. EYRING and G. HALSEY, *Symposium on High Polymer Physics*, Remsen Press, New York (1948) 61.

## PLASTICITY

**40. Initiation of Plastic Deformation**

Plastic deformation has been defined (see Art. 17) as the irrecoverable deformation characteristic of materials existing in the ordered (crystalline) state; it is produced by shear forces causing slip along selected crystallographic planes. Therefore, plastic deformation is essentially a discontinuous process, taking place along discrete planes and extending over a length that is a multiple of the atomic distance (the unit or "quantum" of slip). It first affects crystal planes the orientation of which, with regard to the imposed field of stress, is most favorable for the initiation of slip.

Slip bands within single crystals consisting of clusters of atomic slip planes are spaced at minimum distances of the order of magnitude of 0.1 micron; this spacing increases with increasing temperatures. Hence, the limiting spacing is of the same order of magnitude as the limiting crystallite or block size (see Art. 13) and is equally a function of temperature. Initially one or a few slip bands are formed. As the applied shear stress increases, more slip bands develop, their spacing becoming increasingly uniform.

It has been pointed out in Art. 18 that, since slip in real crystals is not limited to a single atomic plane but involves a number of adjacent atomic planes, a certain amount of disorder is produced by atomic distortion within the glide planes. This disorder is necessarily accompanied by a certain thermal instability resembling that of an unordered atomic arrangement. Glide planes

approach therefore the time- and temperature-sensitive behavior of the crystal boundaries, at least during a certain period after their formation, before a new order is established by place changes of activated particles.

In polycrystalline aggregates, grain boundaries exert a strong restricting influence on the initiation of slip within the grains. The grain boundaries also restrain slip propagation through neighboring crystals. If a crystal grain is favorably oriented with regard to slip and is entirely surrounded by less favorably oriented grains, slip bands will develop only when the applied shear energy is sufficient to force several neighboring grains to deform.

Slip in polycrystalline aggregates is essentially a heterogeneous deformation; although a number of crystals are plastically deformed, the rest of the material remains elastic. The interaction between elastic and plastic deformation results in a gradual transfer of the resistance to the applied forces into the elastic part, which in turn produces further local slip within the most favorably oriented crystals within the elastic regions. Slip initiation thus consists of a series of localized "catastrophic" processes along planes of maximum shear. The direction of atomic slip within the individual crystal will necessarily differ from the over-all direction of the common slip band, as the orientations of the crystals differ. Motion along the common slip bands is thus made up of the components of the individual slip motions within the affected crystals in the direction of this band.

The intersection of the continuous slip bands with the surface of the body are the *glide lines* or *Lueder's lines* observable on the polished surface of certain metals, such as mild steel, in the early stages of plastic deformation. Since it follows the direction of maximum shear, the anisotropic plastic deformation occurs along planes inclined at  $45^\circ$  to the direction of the principal stresses.

The actual field of glide lines may initially be identified with the field of principal shear trajectories in the elastic problem; deviations must, however, be expected, particularly during later stages of deformation, because of the disturbance of the elastic stress field produced by the propagation of plastic glide planes.

Several examples of the fields of anisotropic plastic deformation are presented in Fig. 40·1.

In certain metals, particularly cubic body-centered iron, glide lines are very distinctly visible on the polished surface; in others,

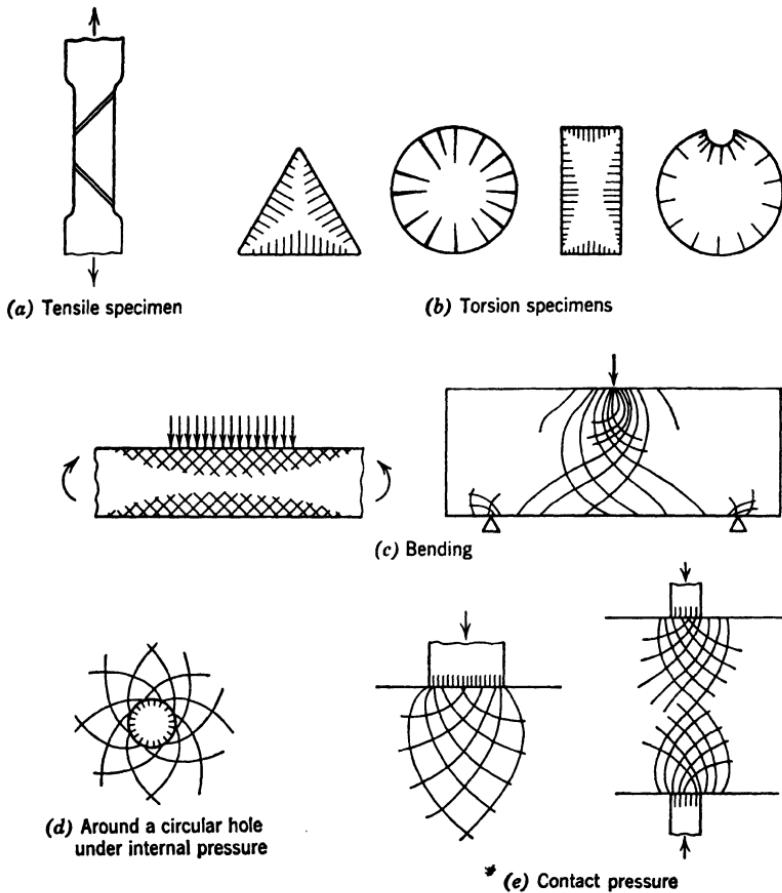


FIG. 40·1 Examples of anisotropic plastic deformation (glide lines).

no anisotropic deformation is visible, and only apparently isotropic areas with a dulled surface appear and gradually spread under increasing load. During the progress of the anisotropic deformation under an applied constant strain rate the measured load fluctuates rapidly about a constant value; the formation of

each new glide line is accompanied by a sudden drop of the load. In tensile tests of mild steel the deformations observed remain within the elastic order of magnitude until the metal in the gage length has been transversed by slip bands (Fig. 40-1a).

Where the distribution of stress is not uniform, as in bending, torsion, or contact pressure, the glide planes associated with initial anisotropic yielding gradually extend from the highly stressed surface into the less stressed interior. In such cases the load observed at a certain strain rate does not fluctuate with the appearance of new slip bands; the anisotropy and heterogeneity of the deformation process is less apparent than it is in homogeneous stress fields with unrestricted propagation of glide lines. The propagation of visible glide lines in nonuniform stress fields is restricted by areas of low stress, causing a relatively early transition from the heterogeneous anisotropic into the statistically isotropic plastic stage; under conditions of homogeneous stress, on the other hand, the heterogeneity and anisotropy of the plastic deformation is very pronounced, and isotropic plasticity develops only under relatively large strains. Moreover, the restriction imposed by the nonuniformity of the stress field on the propagation of the slip bands prevents the spreading of heterogeneous plastic disturbances deep into the elastic region so that stress conditions in this region are not changed. For nonuniform states of stress the initial stage of heterogeneous deformation is therefore less important within the total process of plastic deformation than for homogeneous stress.

Since the fluctuation of the load at the yield limit in the tensile test of mild steel is closely associated with the appearance of glide lines, it is not observed when the initiation of plasticity is gradual and therefore isotropic. Only materials with a sharply defined real yield limit may therefore be expected to show glide lines. This yield limit is represented by the lower level of the fluctuating load during heterogeneous plastic deformation. The upper level of fluctuation is an expression of the delay in the initiation and propagation of glide lines, which is caused by the heterogeneity of the structure of the polycrystalline aggregate, as well as by the time effects associated with the diffusion of foreign atoms into and out of operative slip planes, discussed in Art. 16 and 20.

Visible heterogeneous deformation does not occur under con-

ditions in which the resistance to deformation of the grain boundaries is relatively small. Since slip depends on the resistance of the crystals, which is low, and since a few individual crystals will always be in a position favorable for the development of slip, the material will yield on a macroscopic scale at very low stresses unless it is restrained by the grain boundaries. The occurrence of "catastrophic" slip after local breakdown of the resistance of grain boundaries which produces a sharp yield point, must therefore be associated with highly resistant grain boundaries. Hence, if the manifestation of visible heterogeneous yielding depends on a high resistance to deformation of grain boundaries relative to that of the crystalline areas, this type of deformation will be the less probable, the higher the resistance of the crystalline areas or the lower the resistance of the intercrystalline boundaries.

Some investigators have suggested that the glide lines are a definite property of metals of the body-centered cubic system. Although the evidence is not conclusive enough for this statement to be generally acceptable, it may be assumed that the large number of slip systems of the body-centered crystal reduces the slip resistance of such crystals below that of any other type, thus creating conditions conducive to a sharply pronounced yield limit in polycrystals.

After the over-all deformation has exceeded a certain limit the lines along which slip occurs become so closely spaced that they tend to form a statistically homogeneous area, which appears isotropically plastic. From the boundary of this area further glide lines spread out; it is in this way that the boundary of the plastic region gradually extends in metals with a pronounced yield point. In other metals the anisotropy of slip is limited to the submicroscopic scale; the slip bands are so closely spaced from the very beginning of the deformation that the deformed regions appear isotropic and extend in an apparently isotropic manner.

Hence, in metals with a sharp transition between elastic and plastic deformation, glide lines as well as statistically isotropic plastic areas are simultaneously present. In metals without a pronounced yield point anisotropic deformation takes place on a submicroscopic scale only; phenomenologically the deformation appears homogeneous and isotropic.

The coexistence on a macroscopic scale of heterogeneous anisotropic slip and statistically isotropic plastic deformation requires two different approaches to the solution of problems of plastic deformation. However, heterogeneous slip is usually though not always a transient initial stage, whereas it is the essentially isotropic plastic deformation that is characteristic for plastic behavior. In this stage the development of statistically homogeneous and isotropic plastically deformed regions with a well-defined boundary between the elastic and plastic regions is gradual. The plastic deformation is assumed to spread along the boundary on which the criterion governing the transition from the elastic into the plastic state has been reached.

The only group of problems of plastic deformation in which the effect of the anisotropic yielding is of primary importance and the solution of which cannot, therefore, be based on the assumption of isotropic plastic deformations are the problems of buckling particularly under conditions of 2-dimensional stress. The reason for this is that the limit of instability is associated with the *initiation* of plasticity which, in metals with a sharp yield point, is heterogeneous and anisotropic, and not with the development of an isotropic plastic region at a stage at which the deformation process itself has become instable.

#### 41. Conditions of Plasticity

Although the picture of the gradually spreading isotropic plastic region is an idealization of the real process of transition of the material body from the elastic into the plastic state, this picture provides an adequate approach to the quantitative analysis of many problems of plastic deformation of metals.

In dealing with such problems it is necessary to differentiate between the stage of initiation of plastic deformation, governed by the "yield condition," the deformation throughout the plastic range governed by the "flow condition," and the unloading stage, governed by the "unloading condition." For ideal plastic materials, that is, materials that deform under an unchanging value of the specific resistance to deformation, the "yield condition" and the "flow condition" are identical; that is, the condition of initiation of plastic deformation is satisfied throughout the full range of plastic deformation or flow.

Slip within an individual crystal starts when a critical level of shear energy is applied; this level depends on the orientation of the crystal with regard to the applied stress field, since the deformation of the crystal by slip is anisotropic. If the size of the individual crystal is small in relation to the size of the considered polycrystalline specimen, and if there is no reason to anticipate preferential orientation of crystals within the aggregate, a random distribution of crystal orientations may be assumed to exist as a result of which the initiation of statistically isotropic plastic deformation can be statistically related to the distortional energy producing slip in the individual crystals of the aggregate. Hence, a definite level of distortional energy  $\Phi_y$ , delimits the states of elastic and of plastic deformation, on both the structural and the phenomenological scale. The condition for the initiation of isotropic plastic deformation must therefore have the form,

$$\Phi_0(I_{02}, T) = \Phi_y = \text{const} \quad (41 \cdot 1)$$

where  $\Phi_0$ , the energy of elastic distortion, is a function of the second invariant of the deviator of either stress or elastic strain and of temperature. This is the condition of plasticity introduced independently by Huber<sup>41·1</sup> and v. Mises,<sup>41·2</sup> and redefined by Hencky.<sup>41·3</sup> It will be referred to as the *Huber–Mises–Hencky condition* of plasticity.

Since perfect elasticity is but an ideal limiting condition, and some, no matter how slight, dissipation of the applied energy always takes place above absolute zero temperature if the period of application is long enough (see Art. 19), the transition from the elastic into the plastic state is gradual for all real materials. In some metals, however, the amount of applied energy dissipated below the critical energy level  $\Phi_y$  is so small as to be negligible, whereas the inelastic deformation at a level of applied energy slightly exceeding the critical level is considerable; the transition between the two states is thus rather abrupt, and the setting up of a definite boundary between the elastic and plastic domain has a real physical meaning. In materials with a less sudden transition, the boundary between the elastic and plastic domains has the character of a *designated limit*, being related to the occurrence of an arbitrarily specified amount of irrecoverable deformation, assumed to delimit elastic and plastic behavior.

The condition of initiation of plasticity (eq. 41·1) for constant temperature, can be expressed in terms of principal stresses:

$$\begin{aligned}\Phi_0 &= -\frac{1}{2G} I_{0s2} = \frac{1}{6G} [s_1^2 + s_2^2 + s_3^2 - (s_1s_2 + s_2s_3 + s_3s_1)] \\ &= \frac{1}{12G} [(s_1 - s_2)^2 + (s_2 - s_3)^2 + (s_3 - s_1)^2] = \Phi_y \quad (41 \cdot 2)\end{aligned}$$

Because of eq. 24·31, this condition can also be written in the form:

$$\Phi_0 = \frac{1}{3G} (t_{s1}^2 + t_{s2}^2 + t_{s3}^2) = \Phi_y \quad (41 \cdot 3)$$

$\Phi_y$  is an empirical energy limit which must be derived from experiment. If it is assumed that the yield stress in simple tension is given by  $s_1 = s_0$ , while  $s_2 = s_3 = 0$ , the yield energy limit  $\Phi_y$  is obtained from the distortional energy in terms of the uniaxial yield stress  $s_0$ ,

$$\Phi_0 = -\frac{1}{2G} I_{0s2} = \frac{1}{6} \frac{s_0^2}{G} = \Phi_y \quad (41 \cdot 4)$$

or

$$-3I_{0s2} = s_0^2 \quad (41 \cdot 4a)$$

Hence, the Huber-Mises-Hencky condition of initiation of isotropic plastic deformation (*yield condition*) in principal stresses,

$$2f = (s_1 - s_2)^2 + (s_2 - s_3)^2 + (s_3 - s_1)^2 = 2s_0^2 \quad (41 \cdot 5)$$

or, in principal shear stresses,

$$t_{s1}^2 + t_{s2}^2 + t_{s3}^2 = \frac{1}{2}s_0^2 \quad (41 \cdot 6)$$

In terms of stress components this condition becomes

$$\begin{aligned}2f &= (s_{11} - s_{22})^2 + (s_{22} - s_{33})^2 + (s_{33} - s_{11})^2 \\ &\quad + 6(s_{12}^2 + s_{23}^2 + s_{31}^2) \\ &= 2s_0^2 \quad (41 \cdot 7)\end{aligned}$$

The function  $f$  has been called by v. Mises the *plastic potential*.

If elastic strain is used, the yield condition may be written in the form:

$$\begin{aligned}\Phi_0 &= \frac{G}{12} [(e_1 - e_2)^2 + (e_2 - e_3)^2 + (e_3 - e_1)^2] \\ &= \frac{G}{12} (g_1^2 + g_2^2 + g_3^2) = \Phi_y\end{aligned}\quad (41 \cdot 8)$$

If the yield strain in simple tension is given by  $e_1 = e_0$  while, under conditions of volume-constant distortion,  $e_2 = e_3 = -\frac{1}{2}e_0$ , the yield energy limit  $\Phi_y$  is obtained from the distortional energy in terms of the uniaxial yield strain  $e_0$ :

$$\Phi_0 = -\frac{1}{2} GI_{0e2} = \frac{G}{12} \cdot \frac{9}{2} e_0^2 = \Phi_y \quad (41 \cdot 9)$$

Hence, the yield condition in terms of principal elastic strains,

$$\frac{2}{9}[(e_1 - e_2)^2 + (e_2 - e_3)^2 + (e_3 - e_1)^2] = e_0^2 \quad (41 \cdot 10)$$

in terms of the components of strain,

$$\begin{aligned}\frac{2}{9}[(e_{11} - e_{22})^2 + (e_{22} - e_{33})^2 + (e_{33} - e_{11})^2] \\ + \frac{1}{3}(g_{12}^2 + g_{23}^2 + g_{31}^2) = e_0^2\end{aligned}\quad (41 \cdot 11)$$

In order to express the condition of plasticity in terms of either stress or strain instead of in terms of energy, Hencky<sup>41·4</sup> has introduced the concepts of the *intensity* of stress  $s_r$  and of the *intensity* of strain  $e_r$ , defining these quantities by the expressions:

$$\begin{aligned}s_r &= \frac{\sqrt{2}}{2} [(s_{11} - s_{22})^2 + (s_{22} - s_{33})^2 + (s_{33} - s_{11})^2 \\ &\quad + 6(s_{12}^2 + s_{23}^2 + s_{31}^2)]^{1/2} \\ &= \sqrt{-3I_{0s2}}\end{aligned}\quad (41 \cdot 12)$$

and

$$\begin{aligned}e_r &= \frac{\sqrt{2}}{3} [(e_{11} - e_{22})^2 + (e_{22} - e_{33})^2 + (e_{33} - e_{11})^2 \\ &\quad + \frac{3}{2}(g_{12}^2 + g_{23}^2 + g_{31}^2)]^{1/2}\end{aligned}$$

$$= \frac{2}{\sqrt{3}} \sqrt{-I_{0e2}} \quad (41 \cdot 13)$$

Hence, the yield conditions,

$$s_r = s_0 \quad \text{and} \quad e_r = e_0 \quad (41 \cdot 14)$$

or

$$-I_{0s2} = \frac{1}{3}s_0^2 \quad \text{and} \quad -I_{0e2} = \frac{3}{4}e_0^2 \quad (41 \cdot 15)$$

These conditions can be written in a slightly different form if they are expressed, as suggested by Nadai<sup>41·5</sup> and Ros and Eichinger<sup>41·6</sup> in octahedral shear stresses and strains (see Art. 24).

The octahedral shear stress (eq. 24·34) according to eq. 41·15 is equal to  $\frac{1}{3}s_0\sqrt{2}$ , and the octahedral shear strain (eq. 25·19) according to eq. 41·15 is equal to  $e_0 \cdot \sqrt{2}$ . Hence,

$$t_o = s_r \frac{\sqrt{2}}{3} = \sqrt{\frac{2}{3}} \cdot \sqrt{-I_{0s2}} = \frac{s_0 \sqrt{2}}{3} \quad (41 \cdot 16)$$

and

$$g_o = e_r \sqrt{2} = \sqrt{\frac{8}{3}} \cdot \sqrt{-I_{0e2}} = e_0 \sqrt{2} \quad (41 \cdot 17)$$

Since both the octahedral stress and strain and the intensities of stress and strain are thus proportional to the roots of the second invariants of the deviators of stress and strain, the relations connecting them,

$$s_r = s_r(e_r) \quad \text{or} \quad t_o = t_o(g_o) \quad (41 \cdot 18)$$

should be identical for volume-constant distortion under any state of stress. Therefore the variables  $s_r$  and  $e_r$  or  $t_o$  and  $g_o$  under complex conditions of stress or strain behave like stress and strain in simple tension or in pure shear. The shape of the functions  $s_r(e_r)$  and  $t_o(g_o)$  can thus be determined by such simple tests.

Within a rectangular system of coordinates of principal stresses ( $s_1, s_2, s_3$ ) the Huber–Mises–Hencky condition of initiation of isotropic plasticity can be represented by a circular cylinder with the radius  $s_0 \sqrt{2/3}$ , the axis of which has equal inclination towards all positive axes of principal stresses. All states of stress within this cylinder are elastic; states defined by points on the surface of the cylinder are those associated with the initiation of plastic flow. Points outside the cylinder are not possible. The circle of intersection of this cylinder with the plane normal

to its axis through the origin of the coordinate axes defined by the equation  $(s_1 + s_2 + s_3) = 0$  represents the yield condition for purely deviatoric states of stress. If the orthogonal stress axes  $s_1, s_2, s_3$  are projected on this plane, they project into axes intersecting at  $120^\circ$ . If the scale of the projection is increased by the factor  $\sqrt{2/3}$ , the deviatoric stress components  $(s_1 - p, s_2 - p, s_3 - p)$  are measured directly on the projected axes. In this representation the yield condition is represented by a circle around the origin of radius  $s_0$ .

The yield condition for two-dimensional states of stress is represented by the curve of intersection of the cylinder with the  $(s_1 s_2)$  plane, which is an ellipse: all points defining states of stress within the ellipse are elastic; points on its circumference define conditions of plasticity.

For plane stress with  $s_3 = 0$ , the equation of the ellipse has the form,<sup>41·2</sup>

$$s_1^2 + s_2^2 - s_1 s_2 = s_0^2 \quad (41 \cdot 19)$$

The axis of the ellipse in the direction  $s_1 = s_2$  has the length  $a = s_0 \sqrt{2}$ ; the axis of the ellipse in the direction  $s_1 = -s_2$  has the length  $b = s_0 \sqrt{2/3}$  since, for  $s_2 = 0$ :  $s_1 = s_0$ , and, for  $s_1 = 0$ :  $s_2 = s_0$ ; on the line  $s_1 = s_2$ , the point of intersection is again  $s_1 = s_2 = s_0$ , whereas, on the line  $s_1 = -s_2$ , the point of intersection is defined by  $s_1 = s_0 \sqrt{1/3}$  and  $s_2 = -s_0 \sqrt{1/3}$ .

For plane strain  $s_3 = \mu(s_1 + s_2)$ ; hence, the equation of the ellipse,

$$s_1^2 + s_2^2 - s_1 s_2 \frac{1 + 2\mu - 2\mu^2}{1 - \mu + \mu^2} = s_0^2 \quad (41 \cdot 20)$$

The long axis  $a$  of this ellipse is a function of Poisson's ratio  $\mu$ ; the small axis remains constant  $b = s_0 \sqrt{2/3}$  (Fig. 41·1). In this representation pure shear is defined by  $s_1 = -s_2$ ; the relation between the yield stress in shear  $k$  and in tension or com-

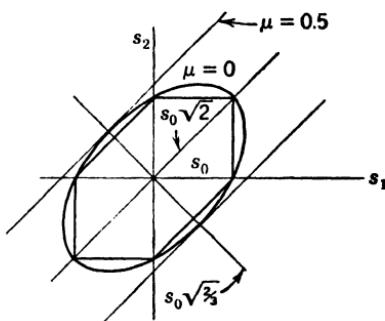


FIG. 41·1 Huber-Mises-Hencky yield condition (ellipse) and Tresca-St. Venant yield condition (hexagon) in two-dimensional representation.

pression  $s_0$  is  $\sqrt{1/3} = 0.577$ . The correctness of this figure for metals has been proved by experiment.<sup>41·7</sup>

With  $k = s_0/\sqrt{3}$  the yield condition (eq. 41·15) can be written in the simple form,

$$-I_{0,2} = k^2 \quad (41\cdot15a)$$

The Huber–Mises–Hencky yield condition based on the criterion of a limiting amount of distortional energy is a relatively late concept. Several alternative conditions of plasticity have been proposed before, and considerable controversy has developed concerning the validity of the different concepts. This controversy was to a certain extent the result of inaccurate terminology, of lack of differentiation between conditions of yielding and conditions of rupture, and of lack of differentiation between conditions governing the initiation of plastic deformation and those governing its progress.

The oldest condition of plasticity is that proposed by Coulomb,<sup>41·8</sup> Tresca,<sup>41·9</sup> and amended by St. Venant,<sup>41·10</sup> which is based on the assumption that in the plastic state the maximum shear stresses have constant values  $s_0/2$ . Hence,

$$(s_1 - s_2)^2 = s_0^2; \quad (s_1 - s_3)^2 = s_0^2; \quad (s_2 - s_3)^2 = s_0^2 \quad (41\cdot21)$$

This system of equations represents in the  $(s_1 s_2 s_3)$  space a hexagonal prism the axis of which has equal inclinations towards all coordinate axes. It is the prism inscribed into the cylinder (eq. 41·5); its intersection with the  $(s_1, s_2)$  plane is the hexagon inscribed into the ellipse representing the Huber–Mises–Hencky condition for plane stress. The relation of 0.5 between the yield limit in pure shear and in homogeneous tension or compression, inherent in the Tresca–St. Venant condition is at variance with most of the experimental results. However, since the discrepancy is the more pronounced, the more gradual the transition between the elastic and the plastic states, it appears that heterogeneous plastic deformation may actually be governed by the Tresca–St. Venant condition.

According to Fig. 41·1 the Tresca–St. Venant condition can be considered a first approximation of the distortion energy condition. Although the difference between those two conditions does not appear to be considerable, it is within the most intensive

states of stress (segments of ellipse) that this difference is largest. In the case of plane, volume-constant strain with  $s_3 = 0.5(s_1 + s_2)$  the distortion-energy condition (eq. 41·5) becomes

$$\frac{1}{2}(s_1 - s_2) = \left[ \left( \frac{s_{11} - s_{22}}{2} \right)^2 + s_{12}^2 \right]^{\frac{1}{2}} = + \frac{s_0}{\sqrt{3}} \quad (41 \cdot 22)$$

and thus differs from the Tresca-St. Venant condition,

$$\frac{1}{2}(s_1 - s_2) = \frac{s_0}{2} \quad (41 \cdot 23)$$

only by the factor  $2/\sqrt{3} = 1.15$ . In the case of plane stress, however, the distortion-energy condition is given by eq. 41·19, while the maximum shear condition retains the form 41·23.

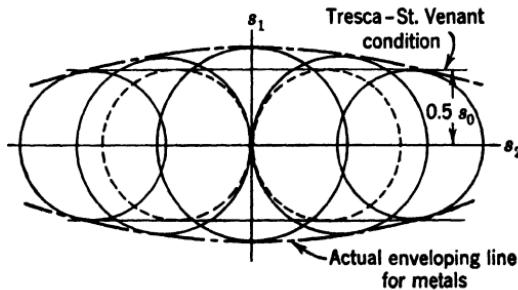


FIG. 41·2 Representation of yield condition as enveloping line of Mohr's stress circles.

The Tresca-St. Venant condition can be presented with the aid of Mohr's circles (Fig. 41·2). However, it has been found by experiment that the parallel enveloping lines at a distance  $t = \pm s_0/2$  which, according to this condition, should delimit the state of plastic flow do not exist.<sup>41·11</sup> Therefore, it cannot be the maximum difference between two of the principal stresses, but a function embodying all three stresses such as the invariant  $I_{0s2}$  which defines the condition of insetting plastic deformation.

Yield tests on nonmetallic substances are rather difficult to interpret because of the arbitrariness of the definition of the "initiation" of inelastic deformation and because of the very pronounced effect of the rate of loading. In general, however, it has been found that the Huber-Mises-Hencky condition of

plasticity is not valid for nonmetallic substances. The Tresca-St. Venant condition of constant shear stress, on the other hand, can be generalized so as to account for the fact that for many nonmetals the critical value of the maximum shear stress which delimits conditions of inelastic deformation is not constant, but a function of the hydrostatic stress  $p$ . This is the Mohr-Kármán hypothesis  $t_{\max} = f(p)$ ,<sup>41·12</sup> where  $f(p)$  is an empirical function of the hydrostatic stress; this function is usually represented as the enveloping curve of Mohr's circles of limiting shear stress (Fig. 41·3). The validity of this condition has been

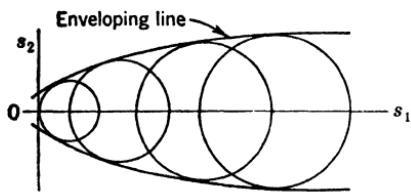


FIG. 41·3 Mohr-Kármán yield condition as enveloping line of circles of principal shear stress.

established for stone, cast iron, and many other porous materials. It is, however, a criterion of rupture in shear rather than one of plastic deformation; this point is frequently overlooked.

For certain natural and synthetic resins the Huber-Hencky-Mises theory has

been found to reproduce the observed behavior fairly well; however, a definite influence of the state of hydrostatic compression or tension has to be introduced. For such materials eq. 41·1 can be written in the form,

$$\Phi_0(I_{02}, T) = f(I_1) \quad (41 \cdot 24)$$

where  $f(I_1)$  is a function of the first invariant, to be determined by experiment. This condition has originally been formulated by v. Mises<sup>41·2</sup> and more recently reintroduced as a "parabolic" yield limit;<sup>41·13</sup> it leads to solutions of a form similar to those of the theory of compressible fluids.

The conditions of constant distortional energy and of constant maximum shear stress are essentially *yield conditions*, governing the *initiation* of plastic deformation. They cannot be expected to govern the *progress* of plastic deformation after its initiation, since this phenomenon is affected by changes in the internal structure of the material taking place during the process. Only under the assumption that in the course of plastic deformation identifiable structural changes do not occur can it be assumed that the yield condition will also describe the process of plastic

deformation after its initiation and may therefore be used as a flow condition. The validity of this assumption is limited to conditions of very small ideally plastic deformation and to conditions of stationary plastic flow of such small velocity that the process may be considered a sequence of states of equilibrium.

## 42. Theories of Plastic Deformation

In dealing with the relations between the mechanical variables governing plastic deformation, the problems are usually divided into two groups. The theories of the first group, known as *theories of plastic flow*, are based on the realization that the calculation of stresses and strains in any general problem of plasticity involves a close analysis of the history of the deformation from the moment of initiation of plastic flow. Hence, a process of plastic deformation is mathematically to be considered as a succession of infinitesimal increments of distortion, and the relation between the mechanical variables can only be a relation between the momentary stress and the *increment* of plastic strain or the *velocity strain*; the strain increments or velocity strains,  $de_i = \dot{e}_i$ ,  $de_{ii} = \dot{e}_{ii}$ , and  $dg_{ij} = \dot{g}_{ij}$ , are defined with reference to the deformed shape of the material body just before the increments have been imposed. The theories of the second group known as *theories of plastic deformation* are based on the simplifying assumption that a step-by-step integration over the entire history of plastic deformation can closely enough be approximated by an averaging process over this deformation history, by which the relations between stress and strain are re-established.

Under certain conditions this averaging process of obtaining the plastic strain from the velocity strain by simple multiplication by the sum of the intervals or the time is more than a first approximation; it represents the rigorous procedure. These conditions require that during the entire deformation period the principal axes of the resultant strain tensor, that is, the sum of the elastic and plastic strain tensors, coincide with the stress tensor; that the ratios of principal stress,  $s_3/s_1$  and  $s_2/s_1$ , remain constant and the stress field not rotate; and, finally, that the density of the body not change. These conditions are automatically fulfilled when in the ideally plastic substance in which the elastic deformations are neglected the stress components

remain at their constant value and the plastic strains increase by small amount under conditions of steady flow. Under conditions of combined elastic-plastic deformation, however, when a definite boundary between the elastic and the plastic regions exist which gradually changes in location and shape as the external forces are increased, the field of elastic stresses is continually displaced within the body. Thus, in the case of elastic-plastic deformation the ratios between the principal stresses will not remain constant, and theories of plastic flow should be used generally in order to obtain rigorous solutions. However, the step-by-step integration required in the theories of plastic flow makes the rigorous solution of any problem, even of the relatively simple ones of rotational symmetry such as the thick-walled cylinder under internal pressure, difficult and cumbersome;<sup>42.1</sup> it is doubtful moreover whether the accuracy achieved compared to the approximate solutions by the theory of plastic deformation justifies the effort involved. Thus, until further development of the theory of plastic flow provides more effective tools for the solutions of problems of plasticity, and until it also becomes evident that the approximate solutions of the problems under investigation are so inaccurate that they cannot be trusted even as approximations, theories of plastic deformation must necessarily remain the principal tool in the solution of technically important problems of plasticity.

The general description of plastic deformation of a material body requires the introduction of two different sets of stress-strain relations: one to be used for *loading*, that is, for conditions of stress that produce an increase of the irrecoverable strain; the other for *unloading*, that is, for conditions affecting the elastic strain only. Since the variation of the work done by the internal forces, both below and beyond the elastic limit, is directly proportional to the variation of the stress or strain intensities, these intensities may be used to define the process of loading and unloading. Thus, a process of loading at a certain point of the considered body is defined as such a change of the state of stress as a result of which either  $(-I_{0s2})$  or the intensity  $s_r$  increases; conversely, for a process to be defined as unloading the invariant  $(-I_{0s2})$  or the stress intensity,  $s_r$ , must decrease. An infinitesimal change of stress for which  $dI_{0s2} = 0$  or  $ds_r = 0$  has been called by Prager a *neutral* change of stress. Any change of

stress that affects the hydrostatic stress only is necessarily neutral.

Since any given neutral change of stress thus differs by an arbitrarily small amount from changes that constitute loading and changes that constitute unloading, it should be required that the stress-strain laws for loading and unloading satisfy this condition of continuity for any neutral change of stress. It has been shown by Prager and coworkers<sup>42-2</sup> that no stress-strain law representing a theory of plastic deformation can satisfy the continuity condition and that only theories of plastic flow can provide the continuity between loading and unloading.

Theories of plastic deformation and theories of plastic flow, in general, have been modeled on the theory of the elastic body and on the theory of the viscous liquid, respectively. Hence, in formulating laws of plasticity the assumptions are usually made that proportionality exists between the stress deviator and either the deviator of strain or the deviator of velocity strain and that for conditions of unloading the stress-strain law is that of linear elasticity. Hence, for loading, under conditions of flow,

$$\mathbf{T}_{0jk} = \lambda \mathbf{E}_{0jk} \quad (42 \cdot 1)$$

where  $\lambda$  is a positive scalar factor of proportionality; for unloading,

$$\mathbf{T}_{0jk} = 2G\mathbf{E}_{0jk} \quad (42 \cdot 2)$$

These assumptions are equivalent with the introduction by v. Mises of a *plastic potential* (see eqs. 41·5 and 41·7) from which the relation in the plastic range between the components of deviatoric stress and of distortional strain (or strain increment) can be derived in the same manner in which these relations in the elastic range are derived from the elastic potential. Hence, with the plastic potential, according to eq. 41·7,

$$de_{0ii} = \frac{1}{\lambda_0} \cdot \frac{\partial f}{\partial s_{ii}} \quad \text{and} \quad dg_{ij} = \frac{1}{\lambda_0} \cdot \frac{\partial f}{\partial s_{ij}} \quad (42 \cdot 3)$$

where  $\lambda_0$  denotes a scalar factor of proportionality. With

$$\frac{\partial f}{\partial s_{11}} = (2s_{11} - s_{22} - s_{33}) = 3(s_{11} - p)$$

and  $\frac{\partial f}{\partial s_{12}} = \frac{\partial f}{\partial s_{21}} = 3s_{12}$  (42 · 4)

and similar equations for the other coordinate axes or, for principal stresses,

$$\frac{\partial f}{\partial s_i} = 3(s_i - p) \quad (42 \cdot 5)$$

the Mises stress-strain or velocity-strain relations are

$$\lambda_0 de_{011} = 3(s_{11} - p) \quad \text{and} \quad \lambda_0 dg_{12} = 3s_{12} \quad (42 \cdot 6)$$

with similar equations for the other coordinate axes; for the principal axes  $i$  of stress and strain,

$$\lambda_0 de_{0i} = 3(s_i - p) \quad (42 \cdot 7)$$

By introducing  $\lambda_0 = 3\lambda$ , eqs. 42·6 and 42·7 can be written in the form,

$$\frac{s_{ii} - p}{d(e_{ii} - e_v)} = \frac{s_i - p}{d(e_i - e_v)} = \frac{s_{ij}}{dg_{ij}} = \text{const} = \lambda \quad (42 \cdot 8)$$

or

$$\frac{s_1 - s_2}{de_1 - de_2} = \frac{s_2 - s_3}{de_2 - de_3} = \frac{s_3 - s_1}{de_3 - de_1} = \lambda \quad (42 \cdot 9)$$

Eq. 42·9 can also be written in the form,

$$\frac{s_1 - s_2}{s_3 - s_1} = \frac{de_1 - de_2}{de_3 - de_1} \quad \text{and} \quad \frac{s_2 - s_3}{s_3 - s_1} = \frac{de_2 - de_3}{de_3 - de_1} \quad (42 \cdot 9a)$$

from which, by subtraction of the two left sides and the two right sides,

$$\frac{3(s_2 - p)}{s_1 - s_3} = \frac{2s_2 - s_1 - s_3}{s_1 - s_3} = \frac{3(de_2 - de_v)}{de_1 - de_3} = \frac{2de_2 - de_1 - de_3}{de_3 - de_1} \quad (42 \cdot 9b)$$

Equation 42·9 states that the Mohr circles for the stresses and the strain increments are geometrically similar, since, no matter what the magnitude of the differences  $(s_1 - s_3)$  and  $(de_1 - de_3)$ , the relations between the radii of the circles determined by the intermediate stress  $s_2$  and the intermediate strain increment  $de_2$  are identical. It is evident, therefore, that in dealing with problems of plasticity strains can be used instead of the strain increments only if during the entire process of deformation the Mohr circles for stresses and strains remain geometrically similar.

With  $\lambda = 2G$ , eqs. 42·8 and 42·9 are transformed into equations of linear elasticity if the strain increments are replaced by strains.

The factor of proportionality  $\lambda$  is obtained from the yield or flow condition, eq. 41·15 if in the expression for the second invariant  $I_{0e2}$  the components of stress are replaced by the components of the strain increment according to eq. 42·7. Thus,

$$\lambda = k/\sqrt{-I_{0e2}} = \frac{s_0}{\sqrt{3}}/\sqrt{-I_{0e2}} = \frac{2s_0}{3de_r} \quad (42 \cdot 10)$$

Hence, the relations 42·6:

$$d(e_{ii} - e_v) = \frac{3de_r}{2s_0} \cdot (s_{ii} - p) \quad \text{and} \quad dg_{ij} = \frac{3de_r}{2s_0} \cdot s_{ij} \quad (42 \cdot 11)$$

For unloading defined by  $de_r < 0$  or for straining for which  $e_r < e_0$ , eqs. 42·11 are transformed into the relations of elasticity:

$$2G(e_{ii} - e_v) = (s_{ii} - p) \quad \text{and} \quad 2G \cdot g_{ij} = s_{ij} \quad (42 \cdot 12)$$

A method of derivation of the stress-strain relations for plastic deformation based on a variational principle introduced by Haar and v. Kármán<sup>42·3</sup> has been suggested by Hencky.<sup>42·4</sup> Haar and v. Kármán attempted to derive the equations governing the state of elastic-plastic equilibrium from the condition that this equilibrium is defined by a minimum value of elastic deformational energy under the auxiliary conditions provided by the three equations of static equilibrium and the yield condition (eq. 41·7). This variational problem results in stress-strain relations of the form,

$$e_{ii} = \frac{1 + \phi}{2G} \left( s_{ii} - \frac{\phi + \frac{3\mu}{1 + \mu}}{\phi + 1} p \right) \quad \text{and} \quad g_{ij} = \frac{1 + \phi}{2G} s_{ij} \quad (42 \cdot 13)$$

where  $\phi$  is a function of the coordinates. The volumetric strain

$$\begin{aligned} e_v &= \sum_1^3 e_{ii} = \frac{1 + \phi}{2G} 3p \left( 1 - \frac{\phi + \frac{3\mu}{1 + \mu}}{\phi + 1} \right) \\ &= \frac{1 - 2\mu}{2G(1 + \mu)} p = \frac{1}{K} p \end{aligned} \quad (42 \cdot 14)$$

Equation 42·14 shows that the relation between  $e_v$  and  $p$  is the same in the elastic and in the plastic range; this was to be expected since inelastic deformation affects the change of shape only and not the change of volume. Equations 42·13 may be written in the form,

$$(e_{ii} - e_v) = \frac{1 + \phi}{2G} (s_{ii} - p) \quad \text{and} \quad g_{ij} = \frac{1 + \phi}{2G} s_{ij} \quad (42\cdot15)$$

or, splitting the elastic and plastic parts of the deformation,

$$(e_{ii} - e_v) = \frac{1}{2G} (s_{ii} - p) + \frac{\phi}{2G} (s_{ii} - p) \quad (42\cdot16)$$

$$g_{ij} = \frac{1}{2G} s_{ij} + \frac{\phi}{2G} s_{ij} \quad (42\cdot17)$$

Whereas the Mises equations describe plastic flow, the Hencky equations refer to elastic-plastic deformation.

The eqs. 42·13 make it clear that, whereas the deformation within the elastic range is governed by the shear modulus  $G$ , the material appears to soften in the elastic-plastic range, the deformation of which is governed by the factor  $(1 + \phi)/G > 1/G$ . The function  $\phi$  has positive values within the plastic area. It is like  $\lambda$  a function of the coordinates and may be a function of time. The equation  $\phi = 0$  is the equation of the surface which delimits the elastic and the plastic domains.

If in flow theories the elastic deformations accompanying plastic flow are taken into account, the superposition of Hooke's law and of eq. 42·11 results in the relation,

$$(\dot{e}_{ii} - \dot{e}_v) = \frac{1}{2G} (\dot{s}_{22} - \dot{p}) + \frac{1}{\lambda} (s_{ii} - p) \quad (42\cdot18)$$

$$\dot{g}_{ij} = \frac{1}{2G} \dot{s}_{ij} + \frac{1}{\lambda} s_{ij}$$

These relations have been proposed by Prandtl and Reuss.<sup>42·5</sup>

The similarity of the eqs. 42·18 with the equations of a Maxwell body is purely superficial; the form of the function  $\lambda$  is such that the equations are homogeneous in time. This becomes

clear when the relation between the equations of plastic deformation 42·17 and those of plastic flow 42·18 are established by differentiating eq. 42·17 with respect to time:

$$\begin{aligned}(\dot{e}_{ii} - \dot{e}_v) &= \frac{1}{2G} (\dot{s}_{ii} - \dot{p}) + \frac{\phi}{2G} (\dot{s}_{ii} - \dot{p}) + \frac{\phi}{2G} (s_{ii} - p) \\ \dot{g}_{ij} &= \frac{1}{2G} \dot{s}_{ij} + \frac{\phi}{2G} \dot{s}_{ij} + \frac{\phi}{2G} s_{ij}\end{aligned}\quad (42\cdot19)$$

The difference between the flow eqs. 42·19 derived from Hencky's equations of plastic deformation, and the Prandtl-Reuss equations of plastic flow 42·18 is in the existence on the right side of the former of the second (plastic) term relating the deviator of velocity strain to that of the stress velocity. Comparison of the third term shows that  $1/\lambda = \phi/2G$  and that therefore eqs. 42·18 are homogeneous in time.

If the proportionality between the deviator of stress and that of velocity strain is restricted to states of stress above the yield limit, whereas below this limit the material is considered undeformable, its behavior is essentially that of a viscous body with a yield limit. In the theory of ideal bodies such a material is known as a "Bingham body" or a *riscoplastic* material.

Within the so-defined material the assumption of proportionality between the velocity strain and the instantaneous value of stress is qualified by the condition that at the yield limit  $F(I_{02}) = \Phi_0(I_{02}) - \Phi_y = 0$  all velocity strains vanish. It is therefore necessary to replace, as the relevant dynamical variable, the tensor or deviator of stress by the so-called tensor or deviator of *overstress*, which is fully defined by the existing state of stress and vanishes at the yield limit. This procedure has been suggested by Fromm<sup>42·6</sup> and developed by Prager.<sup>42·7</sup>

By denoting by  $s'_1$ ,  $s'_2$ , and  $s'_3$  the projections, on the axes of principal stress  $s_1$ ,  $s_2$ , and  $s_3$ , of the vectors representing the tensor of overstress at a point  $P$  situated beyond the yield limit  $F(s_1s_2s_3) = 0$  within the  $(s_1s_2s_3)$  space, a function  $H(s_1s_2s_3) = \text{const}$  can be defined, which describes a single-parameter family of surfaces such that the vector associated with a point  $P$  is normal to the surface passing through  $P$ . The components of a normal to the surface are proportional to  $\partial H/\partial s_1$ ,  $\partial H/\partial s_2$ , and

$\partial H / \partial s_3$ ; thus, the normal (unit) vector has the directional cosines,

$$l_1 = \frac{1}{K} \cdot \frac{\partial H}{\partial s_1}; \quad l_2 = \frac{1}{K} \cdot \frac{\partial H}{\partial s_2}; \quad l_3 = \frac{1}{K} \cdot \frac{\partial H}{\partial s_3} \quad (42 \cdot 20)$$

where, because of  $l_1^2 + l_2^2 + l_3^2 = 1$ ,

$$K = \left[ \left( \frac{\partial H}{\partial s_1} \right)^2 + \left( \frac{\partial H}{\partial s_2} \right)^2 + \left( \frac{\partial H}{\partial s_3} \right)^2 \right]^{\frac{1}{2}} \quad (42 \cdot 21)$$

If the intensity of the vector, the components of which are  $s'_1, s'_2, s'_3$ , is defined by the function,

$$G = \sqrt{s'_1'^2 + s'_2'^2 + s'_3'^2} = G(s_1, s_2, s_3) \quad (42 \cdot 22)$$

the principal stress components of the tensor of overstress are given by the expressions:

$$s'_1 = \frac{G}{K} \cdot \frac{\partial H}{\partial s_1}; \quad s'_2 = \frac{G}{K} \cdot \frac{\partial H}{\partial s_2}; \quad s'_3 = \frac{G}{K} \cdot \frac{\partial H}{\partial s_3} \quad (42 \cdot 23)$$

Since the overstress tensor vanishes at the boundary between the elastic and plastic region, the function  $G$  must be zero for  $F = 0$ . The yield limit is thus equally defined by  $G = 0$ . Under the assumption of isotropy both functions  $G$  and  $H$  are necessarily functions of the invariants of the stress tensor:

$$\begin{aligned} G &= G(I_{1s}, I_{2s}, I_{3s}) \\ H &= H(I_{1s}, I_{2s}, I_{3s}) \end{aligned} \quad (42 \cdot 24)$$

Whereas, as a result of the introduction of the Huber-Mises-Hencky yield condition  $G = F = 0$ , the function  $G = [\Phi_0(I_{0s2}) - \Phi_y]$  depends on the second invariant of the deviator of stress only, this is not *a priori* true of the function  $H$ . Hence, viscoplastic behavior, in general, will be described by higher than second powers of the stress components and thus not only by the second but also by the third invariant.

In order to verify the theoretical assumptions concerning the relations between the mechanical variables in the plastic state of real materials, particularly metals, tests under combined stresses, particularly axial tension and either torsion or internal pressure of thin-walled metal tubes, have been performed.<sup>42·8</sup>

Their results seem to indicate slight systematic deviations from the general relation 42.1 and thus from the conditions 42.8 and 42.9, in which the strain increments are replaced by strains, while supporting in principle the Huber-Mises-Hencky yield condition. If two ratios  $m$  and  $n$  are defined by

$$m = 2 \frac{s_2 - s_3}{s_1 - s_3} \quad \text{and} \quad n = 2 \frac{e_2 - e_3}{e_1 - e_3} \quad (42.25)$$

the validity of the general condition 42.9 would lead to the straight-line relation  $m = n$ . Results of tests<sup>42.8</sup> (Fig. 42.1) show an apparently systematic deviation from this relation. Either this deviation may be due to noncoincidence of the stress and strain tensor in the deformation or to the influence of the overstress tensor in the stress-strain relations by which higher than second-power stress terms are introduced in the yield condition, or it may be ascribed to effects of anisotropy within the material. Since tests under combined stresses are usually performed on thin-walled tubes, the anisotropy inherent in these tubes as a result of the fabrication process must be expected to affect the test results. This anisotropy cannot be eliminated even by the most elaborate heat treatment. Therefore it is not to be expected that test results obtained on such specimen could be in agreement with conditions of deformation derived under the assumption of isotropy of the material.

For orthotropic materials in which the three existing axes of anisotropy intersect at right angles, a modified theory of deformation valid only for rotationless deformations the principal axes of which coincide with the axes of anisotropy, might be established by developing an expression for the distortional energy of the orthotropic body, which would replace the expression  $\Phi_0(I_{02}, T)$

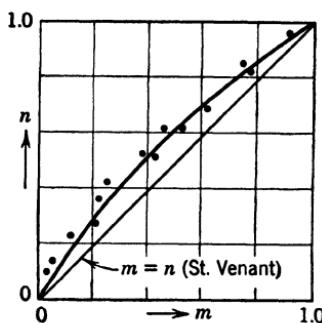


FIG. 42.1 Deviation of Taylor and Quinney's tests<sup>42.9</sup> on metals from the straight-line relation. The curve is derived from Prager's equations of the overstress function  $H = I_{s2} - 0.65(I_{s3}/I_{s2})^2$ .

in the yield condition 41·1. This procedure is open to the objection that the distortional energy of an anisotropic body depends on the relations between the elastic constants. Evidently these constants must appear in the yield condition of the anisotropic material. There is however no reason to assume that they would govern the plastic flow and thus appear in the relations between the stresses and the plastic strain increments. Hence, it appears more reasonable to select, by analogy with the Mises criterion for the plastic flow of an isotropic material, a homogeneous quadratic function of the stress components as the *plastic potential f* of the orthotropic body.

This has recently been done by Hill,<sup>12·9</sup> who introduced the expression,

$$2f = F(s_{22} - s_{33})^2 + G(s_{33} - s_{11})^2 + H(s_{11} - s_{22})^2 + 2Ls_{23}^2 + 2Ms_{31}^2 + 2Ns_{12}^2 \quad (42\cdot26)$$

as the plastic potential of an orthotropic material whose axes of orthotropy coincide with the coordinate axes. If  $s_{01}$ ,  $s_{02}$ ,  $s_{03}$  are the yield stresses in tension in the directions of the axes of orthotropy, and  $k_1$ ,  $k_2$ , and  $k_3$  are the yield stresses in shear with respect to these axes, the constants of the plastic potential are defined by the relations:

$$\begin{aligned} G + H &= \frac{1}{s_{01}^2}; \quad 2F = \frac{1}{s_{02}^2} + \frac{1}{s_{03}^2} - \frac{1}{s_{01}^2}; \quad 2L = \frac{1}{k_1^2} \\ H + F &= \frac{1}{s_{02}^2}; \quad 2G = \frac{1}{s_{03}^2} + \frac{1}{s_{01}^2} - \frac{1}{s_{02}^2}; \quad 2M = \frac{1}{k_2^2} \\ F + G &= \frac{1}{s_{03}^2}; \quad 2H = \frac{1}{s_{01}^2} + \frac{1}{s_{02}^2} + \frac{1}{s_{03}^2}; \quad 2N = \frac{1}{k_3^2} \end{aligned} \quad (42\cdot27)$$

By applying eqs. 42·3 to the yield criterion 42·26 the following relations between the components of the stress tensor and the tensor of strain increment are obtained:

$$\begin{aligned} \lambda_0 de_{11} &= H(s_{11} - s_{22}) + G(s_{11} - s_{33}); & \lambda_0 dg_{23} &= Ls_{23} \\ \lambda_0 de_{22} &= F(s_{22} - s_{33}) + H(s_{22} - s_{11}); & \lambda_0 dg_{31} &= Ms_{31} \quad (42\cdot28) \\ \lambda_0 de_{33} &= G(s_{33} - s_{11}) + F(s_{33} - s_{22}); & \lambda_0 dg_{12} &= Ns_{12} \end{aligned}$$

Since  $e_{11} + e_{22} + e_{33} = 0$ , the selection of the plastic potential (eq. 42·26) implies that the superposition of hydrostatic pressure does not affect the distortion of the body and thus its plastic yielding. This assumption is, however, justified only as a first approximation (see Art. 21).

A similar attempt to develop a condition of orthotropic flow has been made by Jackson,<sup>42·10</sup> who succeeded in reproducing test results obtained on steel tubes under combined stress, for which the isotropic theory proved inadequate.

### 43. Residual Stresses following Plastic Deformation

It has been assumed so far that the behavior of the plastic body depends on the instantaneous values of the deviators of stress and of strain or of the integral over the strain increments during loading, but not on previous strain history. However, if a continuous elastic region exists simultaneously with plastic domains, the mechanical behavior of the body will depend on the sequence of load application and unloading. As a result of the irrecoverable deformation produced within the plastic area on loading, a system of residual stresses is introduced on unloading. The residual stresses evidently influence the deformation produced by subsequent loads. This interaction of elastic and plastic regions creates a *strain-hardening* effect, which although not identical with the work hardening of the material produced by textural microstresses is basically the same phenomenon. The difference is only in the order of magnitude of the inelastic regions involved.

If, under the assumption of ideal plasticity, no elastic area exists in the deformed body, or if the elastic deformations are neglected in comparison with the plastic deformations, the stresses in the ideal plastic body are determined by the initial yield condition and remain invariable during the progress of plastic deformation. However, unless a relation between the deviators of stress and of strain velocity is introduced and a problem of viscous flow formulated, the deformations remain indeterminate, although the stress problem is determinate.

The deformational behavior of a material in the range of elastic-plastic equilibrium, and the transition from conditions of elastically restrained or "*contained*" plastic deformation to a state of unrestrained flow under determinate stresses can be

illustrated by a very simple structural model proposed by Hencky<sup>43.1</sup> and shown in Fig. 43·1. Denoting the cross section of the members  $S_1$  and  $S_2$  by  $A_1$  and  $A_2$ , respectively, the analysis of the statically indeterminate system under the action of the force  $P$  gives the relations:

$$S_2 = \frac{P}{1 + 2 \frac{A_1}{A_2} \sin^3 \alpha} \quad \text{and} \quad S_1 = S_2 \frac{A_1}{A_2} \sin^2 \alpha \quad (43 \cdot 1)$$

The elastic energy  $W$  of the system is expressed by

$$\begin{aligned} W &= \sum^n S_i^2 \frac{l_i}{EA_i} = \frac{2h}{EA_1 \sin \alpha} S_1^2 + \frac{h}{EA_2} S_2^2 \\ &= k_1^2 S_1^2 + k_2^2 S_2^2 \end{aligned} \quad (43 \cdot 2)$$

Under the specific assumptions  $\alpha = 30^\circ$  and  $A_1/A_2 = 4$ , the forces in the members  $S_1 = S_2 = P/2$ . The upper limit of perfectly elastic condition, defined by  $S_2 = s_0 A_2$ , where  $s_0$  denotes the yield limit, is therefore reached under the load  $P_{el} = 2s_0 A_2$  (Fig. 43·2). The initiation of unrestricted plastic flow is defined by the load  $P_{pl} = 5s_0 \cdot A_2$  under which the members  $S_1$  start to deform plastically. The range  $P_{el} < P < P_{pl}$  is governed by the laws of contained elastic-plastic deformation: the application of any load within this range produces, on unloading, a system of residual stresses affecting the subsequent load cycle. Thus, for instance, if  $P = 4s_0 A_2$ , the elastic forces are  $S_1 = 2s_0 A_2$  and  $S_2 = 2s_0 A_2$ , and the stresses  $s_1 = \frac{1}{2}s_0$ ,  $s_2 = 2s_0$ . Since  $s_2 = s_0$  defines plasticity,  $S_2$  cannot exceed  $s_0 A_2$ , so that the remainder of the load has to be taken on by  $S_1$ ; therefore, the real forces are  $S_1 = 3s_0 A_2$  and  $S_2 = s_0 A_2$ , and the stresses  $s_1 = \frac{3}{4}s_0$ ,  $s_2 = s_0$ . On unloading the member  $S_2$  has become too long to follow the full recovery of strain in  $S_1$ , which would be

$\frac{2S_1 l}{EA_1} = 3s_0 \frac{h}{E}$ ; its own contraction is only  $s_0 \frac{h}{E}$ . Hence, an

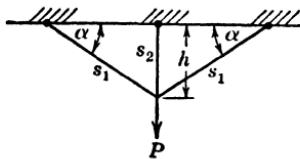


FIG. 43·1 Structural model for illustration of states of elastic-plastic equilibrium and residual stresses.<sup>43·1</sup>

internal force  $X_2$  is necessary to force  $S_2$  back into a compatible position. This force is obtained from the condition,

$$X \left( \frac{h}{A_2 E} + 2 \frac{2h}{4A_2 E} \right) = +2s_0 \frac{h}{E}$$

Hence  $X = s_0 A_2$ , and the residual forces and stresses in  $S_1$  and  $S_2$  are

$$\bar{S}_1 = -\bar{S}_2 = s_0 A_2; \quad \bar{s}_1 = \frac{1}{4}s_0, \quad \text{and} \quad \bar{s}_2 = -s_0$$

Therefore, under a second load application of  $P = 4s_0 A_2$ , the stresses are  $s_1 = \frac{1}{2}s_0 + \frac{1}{4}s_0 = \frac{3}{4}s_0$  and  $s_2 = 2s_0 - s_0 = s_0$ , which is still within the elastic range. By previous plastic deformation the system has thus become elastic for the price of an inelastic displacement  $\Delta$  of magnitude,

$$\Delta = \frac{1}{4}s_0 \frac{2}{E} 2h = \frac{s_0 h}{E}$$

This elasticity is due to the storing up of an energy potential within the system; because this energy neither can be recovered mechanically nor is dissipated, it has been called by G. Taylor<sup>43-2</sup> *latent energy*. The amount of *latent energy*  $W_R$  is, however, only a portion of the total energy  $W$  expended in deforming the structure,

$$W = \frac{s_0 h}{E} 3s_0 A_2 = 3s_0^2 \frac{h A_2}{E}$$

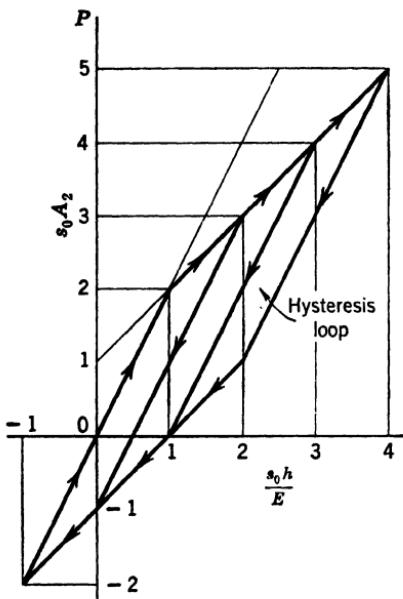
since

$$W_R = \frac{1}{2E} \cdot \frac{h^2}{4} s_0^2 4A_2 + \frac{1}{2E} s_0 2A_2 h = s_0^2 \frac{h A_2}{E} = \frac{1}{3}W$$

If the load is increased beyond  $P = 4s_0 A_2$ , the residual stress in  $S_2$  would exceed  $\bar{s}_2 = -s_0$ , which is not possible. Thus, only the residual stresses pertaining to  $P = 4s_0 A_2$  remain permanently in the system. For loads exceeding  $4s_0 A_2$ , permanent deformation is produced during each load cycle and forcibly reversed during unloading; work is therefore permanently expended and appears as *hysteresis*.

If, after a load  $P_{EI} < P < 4A_2 s_0$  has first been applied and the system made elastic with respect to every repetition of this load, a load is applied in opposite direction, plastic deformation sets

in at a lower load than the elastic limit of the initial (undeformed) system. This so-called *Bauschinger effect* under reversed load is produced by the residual stresses due to the load  $P$  which, by increasing the stresses under the load  $(-P)$  produce an earlier



- (a)  $P = 2s_0A_2$  Elastic
- (b)  $P = 3s_0A_2$  Elastic-plastic: yield point on reversal under  $(-P) = s_0A_2$
- (c)  $P = 4s_0A_2$  Elastic-plastic: yield point on reversal under  $(-P) = 0$
- (d)  $P = 5s_0A_2$  Fully-plastic: yield point on release under  $P = s_0A_2$

FIG. 43.2 Work hardening, Bauschinger effect, and hysteresis in structural model of Fig. 43.1 under load.<sup>43.1</sup>

start of plastic deformation in the member  $S_2$ . Thus, for instance, if  $P = 4A_2s_0$ , plastic deformation will set in under any value of  $(-P) > 0$ , because the residual stress due to the load  $P = -4A_2s_0$  is just at the yield limit  $\bar{s}_2 = -s_0$ . Force-deformation diagrams of the model, under different reversed loads  $P$  are shown in Fig. 43.2. Both *hysteresis* and *Bauschinger effect* are clearly visible.

The discussed model has also been used by Prager<sup>43·3</sup> for a general representation of the ranges of problems of contained plastic deformation and of free plastic flow. If the variables  $x = k_1 S_1$  and  $y = k_2 S_2$  are introduced into eq. 43·2, this equation becomes  $(x^2 + y^2) = r^2 = W$  which is the equation of a

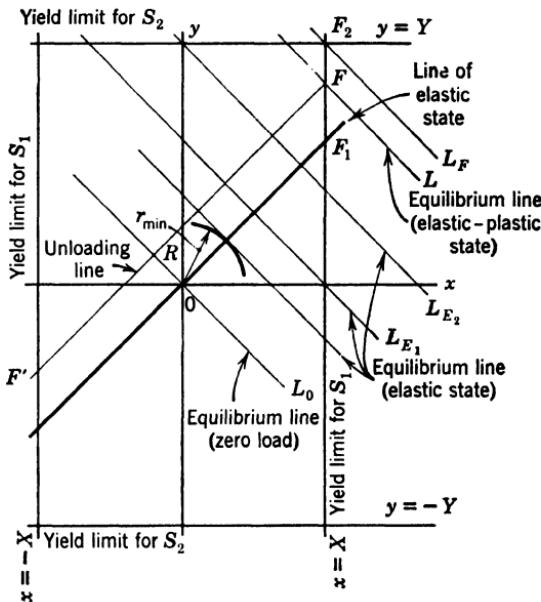


FIG. 43·3 Prager's representation<sup>43·3</sup> of the general elastic-plastic behavior of the Hencky model.

circle. Any possible response of the model to an external load can be represented by a point on the straight line given by the equation of equilibrium of the system:

$$2S_1 \sin \alpha + S_2 = \frac{x}{k_1} 2 \sin \alpha' + \frac{y}{k_2} = P \quad 43·3)$$

The actual elastic response is defined by the point pertaining to the condition of minimum potential energy  $r^2 = W \rightarrow \min$ ; this is the point on the straight line (eq. 43·3), the distance of which from the origin  $x = y = 0$  is a minimum (Fig. 43·3). It is obtained as the point of intersection of the *line of equilibrium*, defined by eq. 43·3, and the *line of the elastic state* which is the normal through the origin to the direction of the parallel lines of

equilibrium. The yield condition for  $S_2$  is given by a straight line parallel to the  $x$  axis at a distance  $y = \pm Y$ , where  $Y = k_2 s_0 A_2$ ; the yield condition for  $S_1$  by a straight line parallel to the  $y$  axis at a distance  $x = \pm X$ , where  $X = k_1 s_0 A_1$ . These straight lines forming the *yield rectangle* delimit the regions of perfect elasticity and of plastic flow; they contain states of elastic-plastic deformation. Thus the conditions defined by points along a line through the origin are elastic between 0 and  $F_1$ , elastic-plastic between  $F_1$  and  $F_2$ , and fully plastic beyond  $F_2$ ; the distance  $(F_2 - F_1)$  defines the range of contained (elastic-plastic) deformation, whereas points beyond  $F_2$  define conditions of free plastic flow.

The family of parallel lines of equilibrium is defined by the parameter  $P$ . The line  $L_0$  through the origin for  $P = 0$  includes all possible states of residual stresses compatible with equilibrium conditions; the line  $L_E$  through  $(0, Y)$  delimits the purely elastic states, the line  $L_F$  through  $(X, Y)$  the states of contained elastic plastic deformation. From this representation it is immediately evident that the minimum condition  $W = r \rightarrow \min$  of the elastic problem must be reinterpreted in the elastic-plastic problem by introducing the auxiliary condition of compatibility of the elastic-plastic state of the structure with the yield condition. Thus the actual response of the model to any load  $P_{\text{el}} < P < P_{\text{fl}}$ , defining an equilibrium line  $L$  between  $L_E$  and  $L_F$ , is represented by the point on  $L$  the distance of which from the origin is the minimum, compatible with the restriction that the point must be located on either of the lines  $x = X$  or  $y = Y$ .

The response to a load cycle  $(0 \rightarrow P \rightarrow 0)$  is represented in the loading stage by the succession of points on the line of elastic state up to  $F_1$  and by points of  $x = X$  up to the maximum load defined by point  $F$ ; in the unloading stage by points on a line through  $F$  parallel to the line of elastic state. The point of intersection  $R$  of this line with the line of residual stresses  $L_0$  defines the state of residual stresses introduced by the load  $P$ . Evidently  $R$  must be located within or on the boundaries of the elastic rectangle. Under reversed direction of the load the limit of purely elastic response is defined by the point of intersection  $F'_1$  of the unloading line of the elastic state  $FR$  with  $x = -X$ . The difference between the length  $RF$  and  $RF'$  represents the *Bauschinger effect* of the model. Being a state of elastic equilib-

rium, the state of residual stresses defined by the point  $R$  on the unloading line must necessarily fulfill the minimum condition  $r \rightarrow \min$ .

The consideration of the model has shown the validity of a minimum principle in the theory of the contained elastic-plastic deformation; this minimum principle, however, is not identical with the principle of minimum strain energy which governs the elastic behavior. By the introduction of the yield condition as an auxiliary condition of the variational problem, the condition of minimum elastic strain energy is changed in the domain of contained elastic-plastic deformation into a condition of minimum distortional energy. Since the elastic distortional energy governs the plastic deformations, a minimum of distortional energy will necessarily be associated with minimum values of the plastic deformations. If plastic resistance is defined in terms of the external forces producing a certain amount of plastic deformation, the minimum condition of distortional energy is equivalent to a condition that has been formulated as a special "condition of maximum plastic resistance," valid within the range of plastic deformation.<sup>43-4</sup> In this range the total strain energy will generally not attain extreme values. Other variational principles have recently been established<sup>43-5</sup> for the plastic body, according to which the stress variation of the plastic work is an absolute maximum, whereas the velocity strain variation is a minimum.

### References

- 41·1 M. T. HUBER, *Czasopismo Techniczne*, Lwow (1904).
- 41·2 R. v. MISES, *Nachrichten Ges. d. Wiss. Goettingen* (1913) 582; *Z. angew. Math. & Mech.* **8** (1928) 161.
- 41·3 H. HENCKY, *Z. angew. Math. & Mech.* **5**, (1925) 116.
- 41·4 H. HENCKY, *Trans ASME* **55** (1933) APM-55-18.
- 41·5 A. NADAI and E. A. DAVIS, *J. Applied Phys.* **8** (1937) 205.
- 41·6 M. Ros and A. EICHINGER, *Ber. Materialprüf. Anst. E.T.H. Zurich* **34** (1927).
- 41·7 F. B. SEELEY and W. PUTNAM, *Univ. Illinois Eng. Expt. Sta. Bull.* **115** (1919).
- 41·8 C. A. COULOMB, *Théorie des machines simples*, Paris (1821).
- 41·9 H. TRESCA, *Mém. présentées par divers savants* **18** (1868) 733.
- 41·10 B. DE ST. VENANT, *Compt. rend.* **70** (1870) 473; **74** (1872) 1009.
- 41·11 W. LODÉ, *Z. Physik* **36** (1926) 913.
- 41·12 O. MOHR, *Z. Ver. deut. Ing.* **44** (1900) 1524; **45** (1900) 740.

41·13 R. v.MISES, *Reissner Anniversary Volume*, J. W. Edwards, Ann Arbor (1949) 427.

42·1 R. HILL, E. H. LEE, and S. J. TUPPER, *Proc. Roy. Soc. A* **191** (1947) 278.

42·2 G. H. HANDELMAN, C. C. LIN, and W. PRAGER, *Quart. Applied Math.* **4** (1947) 397.

42·3 A. HAAR and TH. v.KÁRMÁN, *Nachrichten Ges. d. Wiss. Goettingen* (1909) 204.

42·4 H. HENCKY, *Z. angew. Math. & Mech.* **4** (1924) 323.

42·5 E. REUSS, *Z. angew. Math. & Mech.* **10** (1930) 266.

42·6 H. FROMM, *Ingenieurarchiv* **4** (1933) 439.

42·7 W. PRAGER, *Mém. sci. math.* **87** (1937).

42·8 E. A. DAVIS, *J. Applied Mechanics* **10** (1943) A-187.  
S. J. FRAENKEL, *J. Applied Mechanics* **15** (1948) 193.  
G. I. TAYLOR and H. QUINNEY, *Phil. Trans. Roy. Soc. A* **230** (1931) 323.

42·9 R. HILL, *Proc. Roy. Soc. A* **193** (1948) 281.

42·10 L. R. JACKSON, K. F. SMITH, and W. T. LANKFORD, *Metals Technol.* (1948) T.P. 2440.

43·1 H. HENCKY, *loc. cit.* 42.4.

43·2 G. I. TAYLOR and H. QUINNEY, *Proc. Roy. Soc. A* **143** (1934) 307;  
**163** (1937) 157.

43·3 W. PRAGER, *J. Aeronaut. Sci.* **15** (1948) 337.

43·4 M. A. SADOWSKY, *J. Applied Mechanics* **10** (1943) A-65.

43·5 R. HILL, *J. Applied Mechanics* **16** (1949) A-16.

C H A P T E R  
**8**

## WORK HARDENING OF POLYCRYSTALLINE METALS

### 44. Structural Theories of Work Hardening

When the yield limit of a metal specimen is exceeded, the plastic deformation under either constant or fluctuating stress is more or less rapidly blocked by changes within the structure of the material. These changes result in a gradually increasing resistance to further plastic deformation, increasing indentation hardness and increasing fracture strength, as well as in changes in density, electric conductivity, magnetic properties, and resistance to wear. Whereas the elastic shear modulus remains practically unchanged, slight changes occur in the value of the bulk modulus and of Poisson's ratio.<sup>44-1</sup> This process of changing the mechanical properties of metals by plastic deformation is usually termed *strain hardening* or *work hardening*; for the polycrystalline body the latter term is the more adequate, since the changes within the material may be related to an input of work. A general theory of work hardening can be developed on the basis of this relation.

The rate of change of mechanical and other properties is related to the rate of irrecoverable change in the internal structure of the polycrystal; it depends necessarily on the initial undeformed structure of the material and is different for polycrystalline aggregates of various initial structure. It also depends on the temperature and the speed at which the material is deformed.

Changes in the physical properties by work hardening are

thus related to some modification that has occurred within the configuration of the crystal grains. Since mechanical properties of polycrystalline metals also depend on the nature of the grain boundaries, changes of those properties are necessarily also related to changes in the character and volume of these boundary regions.

When a metal crystal within the polycrystalline aggregate is permanently deformed, slip takes place along a number of the eligible slip planes which are most favorably oriented with respect to the direction of principal shear of the acting stress system. The spacing of the slip planes and the extent of slip along these planes have been found to depend on the applied rate of strain or of loading.<sup>44-2</sup> If the deformation is produced very slowly, the amount of slip on individual planes is extremely small and the number of planes correspondingly large; under such conditions slip on atomic slip planes is rather uniformly distributed and the extent of slip on any plane so small that it may be invisible under the microscope. Above a certain critical strain or loading rate, on the other hand, the spacing of visible slip planes and the extent of slip along individual planes suddenly increases since the visible slip lines are in fact *slip bands* or *glide lines* formed of a cluster of closely spaced atomic slip planes (see Art. 18).<sup>44-3</sup> The critical over-all strain or loading rate at which the finely distributed laminar slip is transformed into one essentially concentrated within clusters of closely spaced slip planes which divide the crystals into glide lamellas depends on the rate at which applied strain energy is dissipated within the intercrystalline regions. If the rate of application of strain energy is higher than its rate of dissipation within this region, extensive breakdown of the intercrystalline cohesion will occur, accompanied by an abrupt change in the distribution of crystalline slip. It is reasonable to assume that the rather uniform restraint imposed on the deformation of a crystal by slip, which is provided by unbroken relaxing grain boundaries, will favor a fine distribution of slip planes and small amounts of slip on individual planes, as such distribution will enable the crystal to deform plastically within the restraint of the relaxing intercrystalline medium. The propagation of extensive slip on widely spaced clusters of slip planes, on the other hand, requires excessive and rapid deformation of the restraining medium which necessarily leads to its local destruction (Fig. 44-1). Finely distributed slip is thus

characteristic of slowly deforming relatively fine-grained polycrystalline aggregates in which the crystals deform by slip within the slowly yielding boundaries. Such deformation cannot be expected to occur in single crystals, the slip of which is unrestrained and therefore rather sudden and extensive, nor in rapidly deformed or coarse-grained polycrystalline aggregates, since the restraints imposed on the slip along widely spaced slip bands are rapidly removed by the local destruction of the cohesion of the

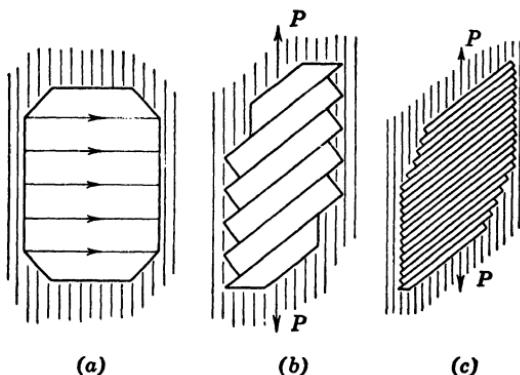


FIG. 44·1 Schematic diagram of extension of crystal (a) in surrounding viscous matrix, (b) by slip along widely distributed glide planes, and (c) by finely distributed slip.

boundary regions. Slip bands in this case are free to propagate through a number of neighboring crystal grains.

Slip is retarded by the distortion of the atomic structure around the slip planes; the amount of distortion created during slip determines the rate of blocking along the slip planes. Therefore, over-all plastic deformation resulting from relatively short slip on many finely distributed slip planes will be less rapidly blocked than plastic deformation resulting from extensive slip along widely spaced clusters of slip planes and associated with the creation and distortion, during the motion, of glide lamella of finite thickness between the slip bands.<sup>44·3</sup>

However, the development of a blocking mechanism by distortion of the atomic arrangement within the slip bands is probably responsible for work hardening only within the range of relatively small plastic deformation. The increase in the resistance to deformation that can be explained by this mecha-

nism alone is usually not of the same order of magnitude as the increase actually observed in polycrystalline aggregates. Thus, the blocking mechanism developing within individual slip bands may account for the work hardening in single crystals; for most metals such work hardening, however, represents only a fraction of the total work hardening observed in polycrystalline aggregates. Therefore, it is probably not so much the blocking of slip *within* the slip bands but the blocking of this slip at or near the grain boundaries which is subsequently overcome by fragmentation, rotation, and elastic distortion and bending of crystal fragments, that produces the observed work-hardening effects of the polycrystal (see Art. 20).

Crystal fragmentation produced in the course of plastic deformation can be made visible by observing the changes in the X-ray diffraction pattern. The broadening of the initially sharp diffraction rings, which indicates a breakup of larger into smaller crystal units, as well as elastic bending and distortion of the fragmented units, the definite limit of this broadening, and the retention of a certain texture of the final X-ray pattern indicate the existence of a limiting size of crystal fragments, to which the initial crystal structure is finally broken down.<sup>44-4</sup> At this limit the thermal instability at the temperature of the deformation, as expressed by the rate of diffusion, becomes so strong that fragments that are smaller than the limiting size tend, if formed, to coalesce rapidly to the limiting or to a larger size.

Crystal fragmentation is probably the principal structural change associated with plastic deformation within the range of strains up to 15 or 20 percent. Changes in the mechanical properties within this range are thus essentially due to the refinement of the initial crystal grains producing crystallites of more or less uniform limiting size and to changes in the volume and energy content of the intercrystalline boundaries.

The primary effect of grain refinement is to increase the specific distortional energy that can be reversibly stored up within an individual crystal grain, since, according to eq. 41·8, the yield limit is proportional to the sum of the squares of the principal shear strains. Assuming that slip within an individual grain of the polycrystal is the result of the energy that produces pure shear in the eligible slip system of the considered crystal, the critical energy level at which slip is initiated along a single one

of the eligible planes is represented by the shear energy that can be released by slip over one atomic distance  $a$ , which is the unit step or *quantum* of slip. Slip over one-half that step only requires energy application; the second half of the step toward the new equilibrium position occurs spontaneously, after the activation energy between the two equilibrium positions has been overcome in the motion over the first half of the unit step. The critical energy level, which depends on the shear strain  $a/2\lambda$ , is, therefore, directly proportional to the square of this ratio where  $\lambda$  is a measure of the linear dimension of the crystal (Fig. 44·2).

The change in the energy content of the boundaries of the broken-up crystal structure is accompanied by an over-all decrease in density of the aggregate of the order of magnitude of 0.1 to 1.0 percent which is a characteristic feature of the work-hardening process of most polycrystalline metals; no measurable decrease in density can usually be found in cold-worked single crystals.<sup>44·5</sup> In the course of the fragmentation of the crystal structure a considerable amount of disorder is created around the glide bands as a result of the distortion of the atomic structure within these bands as well as by rotation, distortion, and bending of the crystal fragments (glide lamellas) during deformation. To stabilize this distorted structure a system of textural micro-stresses of very high potential must be built up during unloading. Thus, a certain percentage of the volume of cold-worked polycrystalline metals will contain particles connected by bonds of exceptionally high energy content; the distances between particles within this volume must markedly exceed those in the remaining crystal volume. The potential energy introduced into and *latently* stored within the polycrystalline structure by the fragmentation of crystals is therefore contained essentially within the small volume of crystal boundaries and distorted slip bands.<sup>44·6</sup> This conclusion has been confirmed by the comparison of changes in the lattice parameter observed in the course of work hardening, with the over-all decrease in density of the material, or with the amount of potential energy latently stored up within the frag-

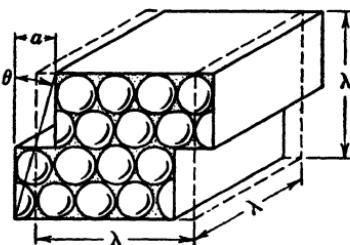


FIG. 44·2 Slip of a crystallite.

mented structure and measured by the difference of the heat content of the undeformed and of the work-hardened structure. The amount of latent energy or the density decrease sufficient to produce an over-all elastic strain consistent with the observed change of the lattice parameter was found to be much smaller than the actually measured latent energy or the change of density. Thus the latent energy and the local density changes must be concentrated mainly within the intercrystalline boundaries and within the distorted slip bands.

At all stages of the fragmentation process, slip within the individual crystals of the polycrystalline aggregate can be effectively blocked by an increased resistance to deformation of the grain boundaries and distorted slip bands, such as that produced by the diffusion of foreign atoms into those regions (*precipitation*). Thus, slip delay and discontinuous yielding, which in hexagonal single crystals is caused by the blocking of the operative slip planes but cannot occur in cubic single crystals because of the large number of alternative slip planes (see Art. 20), can be produced in cubic polycrystals, such as steel and aluminum, by the blocking of the propagation of glide planes across grain boundaries.

A change of deformational properties due only to the refinement of the crystal grains would necessarily be independent of the direction of applied strain, since changes within a statistically isotropic material cannot depend on the direction of the strain. However, the change in the energy content of the distorted atomic layers surrounding the crystal fragments, which results from the system of textural stresses stabilizing the work-hardened structure after fragmentation, introduces a directional effect of a similar type to that discussed in connection with the occurrence of residual stresses in the elastic-plastic deformation of Hencky's structural model (see Art. 43). In the same manner in which the effect of systems of macroresidual stresses is not independent of the direction of the applied force, changes of deformational properties due to the existence of a system of textural stresses depend on the direction of the straining. The expression of this dependence in work-hardened metals is the *Bauschinger effect*; the resistance to plastic deformation as defined by the yield limit is increased in the direction of the work-hardening strain; it is reduced for strains applied in the opposite direction. For

strains in intermediate directions the changes of deformational properties are rather erratic.

When an advanced stage is reached in the process of crystal fragmentation without fracture having occurred, the mechanism of plastic deformation by slip and fragmentation changes gradually and becomes one of formation of a definite texture (orientation), by the rotation and temporary breakup of fragments, which are subsequently re-formed in the direction of the largest strain velocity (see Art. 20). Changes in mechanical and other properties produced during this stage of the deformation are an expression of the developing anisotropy of the material; the structural pattern defined by grain size and by the volume of grain boundaries remains essentially unaffected, and the hardness increase during this stage is relatively small.

#### 45. A General Law of Work Hardening

On the basis of the preceding discussion four different mechanisms may be responsible for the increase of resistance to plastic deformation of a polycrystalline metal aggregate stressed above the yield limit. Although it is frequently attempted to explain the work-hardening phenomenon in terms of a single one of those mechanisms, it is most probably the joint operation of all four mechanisms simultaneously or consecutively that is responsible for the observed work-hardening effect over the whole range of deformations. Work-hardening may thus be the result of

1. The increase in resistance to slip within a single crystal, due to the creation and propagation under applied shear of distortions within the atomic lattice (dislocations).
2. The increase in resistance to plastic deformation of a polycrystalline aggregate produced by the fragmentation of crystals and the rotation, elastic distortion, and bending of crystal fragments. The increased resistance is due primarily to the fact that the specific shear energy required to initiate slip in a single crystal is an inverse function of the square of the crystal size.
3. The stabilization of the fragmented and distorted crystal structure by a system of textural stresses set up during unloading after fragmentation. This system stabilizes the distorted structural pattern by introducing into the material a certain amount of latently stored potential energy.
4. Anisotropic change in resistance to plastic deformation

produced by practically volume-constant deformation associated with rotation, breakup, and rapid re-formation of crystal fragments of limiting size in the direction of the maximum strain velocity.

The mechanism of setting up a system of textural stresses within the deformed structure can be interpreted in two different ways, following two different schools of thought. If it is assumed that the textural stresses are concentrated within the interfaces of the crystal fragments, the mechanism is consistent with the concept of crystal fragmentation; if, on the other hand, it is assumed that the textural stresses are associated with distortions on an atomic scale<sup>45-1</sup> (dislocations), the work-hardening effect may be attributed to an increase in the density of the dislocations produced in the course of the deformation, without the concept of fragmentation being introduced. There is, in fact, no contradiction between the two assumptions if it is recalled that block boundaries can be defined as planes of concentration of dislocations (see Art. 13). Thus, an increase of the density of dislocations can be interpreted as an increase in the extent of the volume of interfaces, which can only result from a reduction of the size of the undistorted crystal regions. Hence, the concept of work hardening due to the potential energy of the increasingly dense dislocations is essentially only a different interpretation of the concept of work hardening due to a fragmentation process; in order to account for the work-hardening limit, which is attained when the structure is made of fragments of limiting size, a limit of maximum density of dislocations must be introduced.

Although the different work-hardening mechanisms are assumed to be jointly operative, the degree of their relative importance probably changes with the initial properties of the material and with the progress of work hardening. Atomic distortion (mechanism 1) is probably responsible for the initial work hardening associated with small plastic deformation, or for the work hardening of polycrystals consisting of a small number of crystals. Beyond the range of small deformations the work-hardening effects resulting from crystal fragmentation and storing up of potential energy, that is, mechanisms 2 and 3, may be assumed to be mainly responsible for producing the observable relations of stress and strain. Reorientation of frag-

ments, which does not produce appreciable hardening, is operative in the formation of texture under large unidirectional strains; therefore in polycrystalline metals the texture-forming mechanism comes into play at a late stage, when crystal fragmentation has been nearly completed and further deformation can only proceed by directional reorientation of the crystal structure under conditions of considerable thermal instability.

A feature common to the mechanisms 1, 2, and 3 will be used to develop a general function of work hardening. This feature is the dependence of the resistance to slip initiation within a crystal or within a polycrystalline aggregate on an inverse function of a geometrical parameter  $\lambda$ , which is considered a measure of the average linear dimension of the crystals or the crystal fragments forming the aggregate.

Considering an elementary crystal cube of linear dimension  $\lambda$  (Fig. 44·2), subject to pure shear, slip is initiated by the gliding of one atomic plane over another over one atomic distance, with accompanying release of stored-up potential energy, when the elastic shear strain  $g$  attains the value  $a/2\lambda$ , that is, when the upper plane is displaced against the base plane by one-half the atomic distance  $a$ . Hence, the elastic energy per unit volume which can be reversibly stored up before slip starts, the so-called *resilience*  $\frac{1}{2}Gg^2 = \text{const } 1/\lambda^2$ ; it is thus inversely proportional to the square of the parameter  $\lambda$ . It is this resilience that provides a measure of the *hardness*  $H$ , by delimiting the energy level at which plastic deformation starts.

If it is assumed that the increase in hardness  $H$  is not or not only due to the resilience of individual crystals increasing with decreasing parameter  $\lambda$  but to the storing up of latent energy within the fragmented structure, the hardness increase  $dH$  would be proportional to the increase, in the course of fragmentation, of the intercrystalline surfaces within the unit volume of the aggregate. If  $m$  crystals of parameter  $\lambda_m$  making up the unit volume of hardness  $H_m$  are broken up into a unit volume of hardness  $H_n$  consisting of  $n$  crystallites of parameter  $\lambda_n$ , and if the ratio of hardness before and after fragmentation  $H_m/H_n$  is assumed to be proportional to the ratio of intercrystalline surfaces  $A_m$  and  $A_n$  per unit volume before and after fragmentation, the hardness ratio can be expressed by

$$\frac{H_n}{H_m} = \frac{n\lambda_n^2}{m\lambda_m^2} = \frac{\lambda_m}{\lambda_n} \quad (45 \cdot 1)$$

since  $(m\lambda_m^3) = (n\lambda_n^3) =$  unit volume. The hardness  $H$  is therefore inversely proportional to the parameter  $\lambda$ . The same relation is obtained from the dislocation theory.<sup>45.2</sup>

Thus, according to any one of the three work-hardening mechanisms the hardness  $H$  defined in terms of the elastic resilience is inversely proportional to either the square of the parameter  $\lambda$  or to  $\lambda$  itself, or to some intermediate power  $\lambda^n$ , where  $1 < n < 2$ . It is on the basis of this conclusion that a general law of work hardening can be developed. This law describes the change of state in terms of the process of energy transformation which accompanies the changes in the structural pattern of the metal. It can therefore be expressed in terms of the basic thermodynamic relation (eq. 28·6); for each structural pattern defined by the energy  $W_D$ , irrecoverably expended in producing it, a limiting amount of distortional energy  $\Phi_0 = H$  exists, up to which  $dW_D/dt = 0$ . The function  $\Phi_0 = H = f(W_D)$  represents the basic work-hardening relation.

Energy being an additive quantity, the hardness of the polycrystalline aggregate can be assumed to be roughly equal to the sum of the individual hardness or resilience values of the constituent crystals; hence,

$$\frac{1}{2} \sum_m^n G g_m^2 V_m = \Phi_0(g) V = \frac{1}{2} G g^2 V \quad (45 \cdot 2)$$

where  $g$  denotes the over-all limiting shear strain and  $V$  the volume of the aggregate, while  $V_m$  and  $g_m$ , respectively, denote the volume and limiting shear strain of an individual crystal of size  $m$  out of the  $n$  sizes of crystals forming the aggregate.

Since the hardness  $H$  of the aggregate is made up of the sum of the hardness  $H_m$  of the constituent crystal sizes, the contribution of the crystal size  $m$  to the hardness of the aggregate will necessarily be  $H_m V_m / V$ . If the structural pattern producing the hardness  $H$  is defined by a certain volume distribution of crystal parameters  $\lambda_m$ , running from  $\lambda_1$  to  $\lambda_n$ , where  $\lambda_1$  denotes the largest crystal size,  $\lambda_n$  the smallest (limiting) size, this pattern is related to the energy  $W_D$  irrecoverably expended in

producing it from the initial structural pattern of the aggregate, defined by the initial hardness  $H_0$ .

The validity of this relation is, however, limited to conditions for which slip, crystal fragmentation, and distortion are the principal operating dissipation mechanisms; this is the case at temperatures at which the thermal stability of the distorted crystal pattern is high and effects of anisotropic reorientation of crystals are slight. At temperatures sufficiently below recrystallization temperature, the energy dissipated by thermal mechanisms will, in general, be negligible. At temperatures near recrystallization temperature, however, the amount of energy dissipated directly into heat without producing a permanent change of structural pattern that would result in increased hardness becomes so large that a definite relation between hardness  $H$  and dissipated energy  $W_D$  no longer exists. Similarly, energy is dissipated without producing appreciable change in hardness, if directional reorientation is the predominant phenomenon during deformation. However, conditions of such thermal instability of the fragmented structure can be expected to become pronounced only at the approach to the limit of hardness. Since for most of the structural metals at room temperatures or even at moderately elevated temperatures fracture will prevent this upper limit of work hardening from being attained, the one-valued relation between hardness  $H$  and dissipated strain energy  $W_D$  may be assumed to hold with fair approximation within the practically important range of deformation.

An increase in hardness  $dH$  is brought about by expending the energy  $dW_D$  to change the volume distribution of crystal sizes through the fragmentation of a part  $v_m$  of the total volume  $V_m$  of crystals of size  $\lambda_m$  and hardness  $H_m$ , into an equal volume of crystals of size  $\lambda_n$  and hardness  $H_n$ . Therefore the rate of work-hardening of the aggregate,

$$\frac{dH}{dW_D} = \sum_1^m (H_n - H_m) \frac{dv_m}{dW_D} \quad (45 \cdot 3)$$

if the considered volume of the aggregate  $V$  is unity; the sum is to be taken over all crystal sizes affected by fragmentation, that is, over all sizes  $m < n$ .

Fragmentation of crystals could be either *gradual*, through

consecutive stages, when every crystal size is broken up into a next smaller size, or *catastrophic*, every crystal size being shattered directly into fragments of limiting size, or it could be of an intermediate nature. It has been found that the first assumption leads to work-hardening functions inconsistent with observations;<sup>16, 3</sup> the catastrophic type of fragmentation is therefore assumed to be prevalent.

If  $q_m$  denotes the ratio of the volume of one crystal of parameter  $\lambda_m$  to the volume of the crystal fragment of limiting parameter  $\lambda_n$ , the ratio of hardness of the two crystal sizes,

$$\frac{H_n}{H_m} = \left( \frac{\lambda_m}{\lambda_n} \right)^2 = q_{m,n}^{-2/3} \quad (45 \cdot 4)$$

if the hardness is defined by the *resilience* of crystal grains alone. If, on the other hand, it is assumed that the increase of hardness is not due to the increasing resilience of the refined crystal grains but results from the *latent energy* stored up in the intercristalline surfaces, the hardness ratio, according to eq. 45·1, is

$$\frac{H_n}{H_m} = \frac{\lambda_m}{\lambda_n} = q_{m,n}^{1/3} \quad (45 \cdot 5)$$

Adopting eq. 45·4 and introducing it into eq. 45·3 give the rate of work-hardening:

$$\frac{dH}{dW_D} = \sum_1^m H_m (q_{m,n}^{-2/3} - 1) \frac{dv_m}{dW_D} \quad (45 \cdot 6)$$

By integration,

$$H = \sum_1^m H_m (q_{m,n}^{-2/3} - 1) v_m + \text{const} \quad (45 \cdot 7)$$

If  $v_m$  denotes the relative volume of grains of parameter  $\lambda_m$  fragmented to date out of the volume  $V_{0m}$  of such crystals initially existing in the unit volume  $V$ , the rate of fragmentation of grains of any particular parameter  $\lambda_m$  may reasonably be assumed proportional to the relative volume of such grains  $V_m = (V_{m0} - v_m)$ , existing at the considered stage of fragmentation defined by  $W_D$ . This assumption is generally used in physical disintegration (decay) processes. Hence,

$$\frac{dv_m}{dW_D} = (V_{m0} - v_m) \frac{1}{\alpha_m} \quad (45 \cdot 8)$$

where  $\alpha_m$  is a factor of proportionality characteristic for the stability of the grain size; its inverse value is proportional to the rate of disintegration. With the condition that  $v_m = 0$  for  $W_D = W_{D0m}$ , integration of eq. 45·8 gives

$$v_m = V_{m0}[1 - e^{-(W_D - W_{D0m})/\alpha_m}] \quad (45 \cdot 9)$$

valid for values of  $W_D > W_{D0m}$ .

The work-hardening curve corresponding to eq. 45·7 has thus different ranges governed by different equations and delimited by the values of  $W_{D0m}$ . In the first range, from  $W_{D01} = 0$  to  $W_{D02}$  the largest size of crystals of parameter  $\lambda_1$  and hardness  $H_1$  is broken up into crystallites of limiting size until the over-all hardness  $H$  of the polycrystalline aggregate (which in the initial stage, that is, for  $W_D = 0$ , is  $H = H_0 = H_1$ ) reaches the hardness  $H_2$  of the grains of parameter  $\lambda_2$ . From there on the fragmentation of the crystal size of parameter  $\lambda_2$  sets in, proceeding simultaneously with the continued breaking up of still existing grains of parameter  $\lambda_1$ , and so on. Thus, for the first stage ( $0 < W_D < W_{D02}$ ), with  $H = H_1$  for  $W_D = 0$ :

$$H = H_1[1 + (q_{1,n}^{\frac{2}{3}} - 1)V_{10}(1 - e^{-W_D/\alpha_1})] \quad (45 \cdot 10)$$

Generally the hardness  $H(W_{Dm})$  within the  $m$ th stage,

$$H(W_{Dm}) = H_1 \left\{ q_{1,m}^{\frac{2}{3}} + \sum_{m=1}^{k=1} [q_{1,k}^{\frac{2}{3}}(q_{k,n}^{\frac{2}{3}} - 1)V_{k0} \right. \\ \left. (e^{-(W_{D0m}-W_{D0k})/\alpha_k} - e^{-(W_{Dm}-W_{D0k})/\alpha_k})] \right\} \quad (45 \cdot 11)$$

where  $q_{1,m}^{\frac{2}{3}} = H_m/H_1$  and  $q_{k,n}^{\frac{2}{3}} = H_n/H_k$ ,  $H_n$  being the hardness of crystallites of limiting size  $\lambda_n$ . The rate of work hardening within this range,

$$\frac{dH(W_{Dm})}{dW_{Dm}} = H_1 \sum_{k=1}^{k=m} \left[ q_{1,k}^{\frac{2}{3}} \cdot (q_{k,n}^{\frac{2}{3}} - 1) \frac{V_{k0}}{\alpha_k} e^{-(W_D - W_{D0k})/\alpha_k} \right] \quad (45 \cdot 12)$$

If the rate of work hardening is plotted against hardness, a straight line,

$$\frac{dH}{dW_D} = C_1 + C_2 H \quad (45 \cdot 13)$$

is obtained if one term only ( $k = 1$ ) exists in eq. 45.12 or, alternately, if  $\alpha_k$  is the same for all crystal sizes. The shape of the function  $H = f(dH/dW_D)$  thus provides an indication of the character of the process of fragmentation.

It is evident that the only change that would be introduced into eqs. 45.10 to 45.13 by adopting a different assumption concerning the operative work-hardening mechanism is to change the powers of all  $q_{k,m}$  and  $q_{k,n}$  ratios in those equations from  $2/3$  to  $1/3$ , as required by eq. 45.5, which would replace eq. 45.4 on the basis of which the work-hardening law has been derived. This change is a change in constants only and does not affect the form of the general work-hardening function.

In applying eq. 45.11 to the interpretation of experimental results for a general state of stress, the proper measure of  $H$  and of  $W_D$  must be introduced. According to eq. 41.4 or 41.9  $\Phi_0$  is proportional to the second invariant of the deviator of stress or of elastic strain. There is, however, no simple measure of  $W_D$ ; it could only be computed directly by integrating the area under the stress-strain diagram (Fig. 29.1). If, in first approximation,  $W_D$  could be assumed to be proportional to the second invariant of the deviator of irrecoverable strain  $I_{0p2}$ , the general work-hardening function  $H = f(W_D)$  would have the invariant form,

$$I_{0e2} = F(I_{0p2}) \quad (45 \cdot 14)$$

In the simplest case of an aggregate with a uniform grain size, eq. 45.10 with  $V_{10} = 1$  governs the entire range of deformation. Hence,

$$H = H_n - (H_n - H_1)e^{-W/\alpha} \quad (45 \cdot 15)$$

and

$$\frac{dH}{dW_D} = \frac{1}{\alpha} (H_n - H) \quad (45 \cdot 16)$$

Relations 45.15 and 45.16 are represented in Fig. 45.1. The slope ( $-1/\alpha$ ) of the relation 45.16 is constant over the entire deformation range. This constancy is an indication that only one crystal size is being broken up.

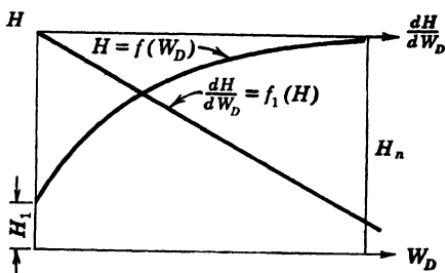


FIG. 45.1 Work-hardening function for fragmentation of single crystal size.

Equation 45.16 can be written in terms of stresses if  $W_D$  is considered to be a function of plastic strain  $\ln p$  alone,

$$\left(\frac{s}{s_\infty}\right)^2 = 1 - \left[1 - \left(\frac{s_1}{s_\infty}\right)^2\right] e^{-(\ln p)^2/\alpha} \quad (45.17)$$

where  $s_\infty$  and  $s_1$  denote the stresses for  $p = \infty$  and  $p = 1$ , respectively. Hence,

$$s = s_\infty \cdot \sqrt{1 - \left[1 - \left(\frac{s_1}{s_\infty}\right)^2\right] e^{-(\ln p)^2/\alpha}} \quad (45.18)$$

The form of this relation for an assumed value  $(s_1/s_\infty) = 0.1$  and for a *stability coefficient*  $\alpha = 1$  is represented in Fig. 45.2. The diagram shows that within the range of logarithmic strain  $0.1 < \ln p < 0.6$  the stress-strain diagram is very nearly straight; it only starts to curve downward at higher values of strain. The apparent straight-line relation is therefore a transient phenomenon and has no physical meaning. Since in tension tests of metal specimens, particularly steel, fracture mostly starts before the curved part of the stress-strain diagram is reached, the apparent straight-line part of the diagram is frequently considered a character-

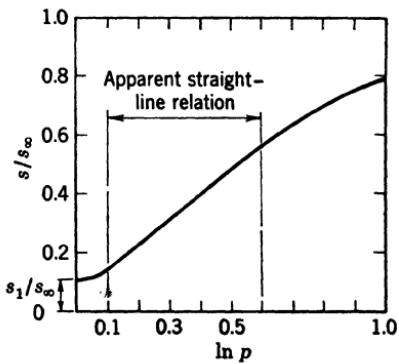
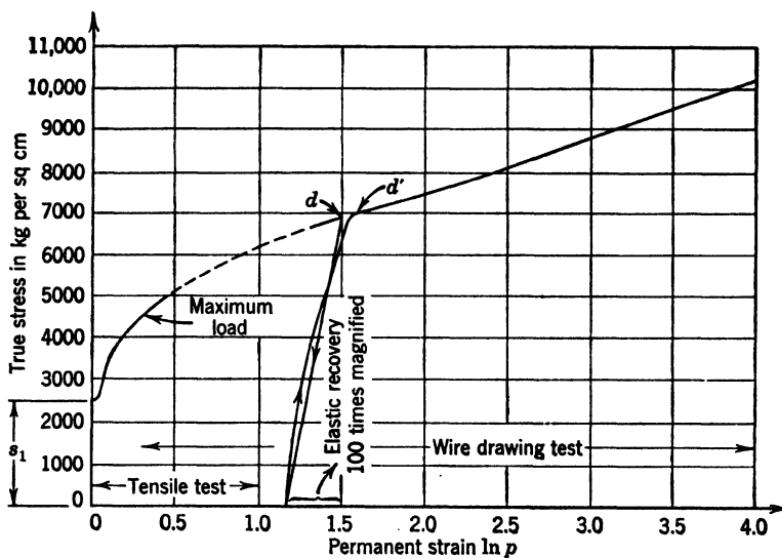
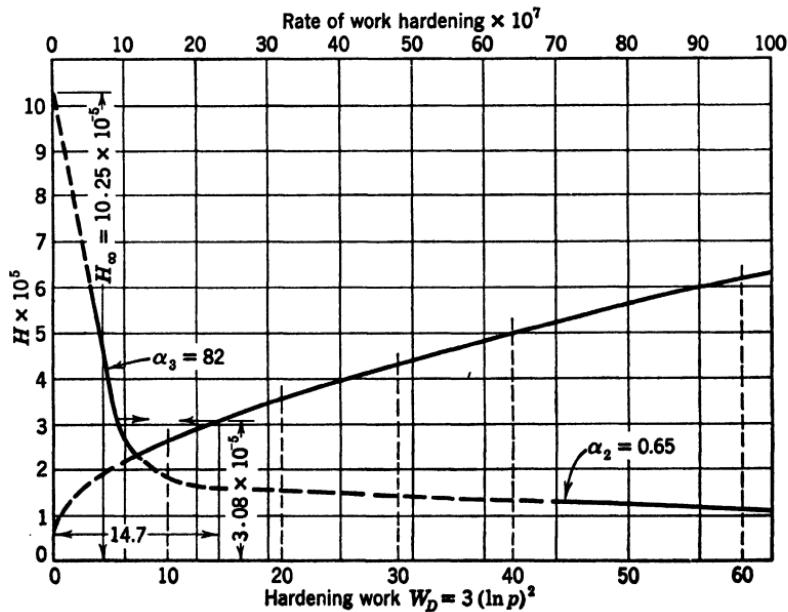


FIG. 45.2 Stress-strain diagram for work-hardening process involving breakup of one crystal size (yield stress  $s_1 = 0.1s_\infty$ ).

FIG. 45.3a Stress-strain curve for mild steel.<sup>45-3</sup>FIG. 45.3b Work-hardening curve for mild steel.<sup>45-3</sup>

istic feature of deformational behavior; there is, however, no justification for this assumption.

Figures 45·3 reproduce the functions  $H = f(W)$  and  $dH/dW_D = f(H)$  in terms of  $(e^2)$  and  $(\ln p)^2$  from a combined tensile-test and wire-drawing experiment performed on mild steel in which very high strain values were reached.<sup>45·3</sup> From the interpretation of the  $dH/dW_D = f(H)$  diagram, it is found that the work-hardening process consists of three distinct stages (of which only two are discernible) in the course of each of which fragmentation of one crystal size takes place; only during the comparatively short curved transition stages does the fragmentation process involve two consecutive crystal sizes. A detailed analysis of the stress-strain curve shows that the true stress  $s = P_{max}/A$  at which necking starts in the tensile test coincides with the stage of deformation when the breakup of the largest crystal size is practically completed. It can be easily verified that necking starts under maximum load when the rate of hardening is no longer able to compensate for the decrease in cross section, by computing the maximum sustained load in the tension test  $P_{max}$  from the minimum condition  $dP = d(sA) = A ds + s dA = 0$  and therefore  $ds/s = -dA/A$  (see Art. 87). The coincidence is therefore to be expected, since the discontinuous decrease in the rate  $dH/H$  and therefore of  $ds/s$  must necessarily produce necking under decreasing load.

#### 46. Time Effects. Thermal Stability of Work Hardening

In the course of the crystal-fragmentation process regions of high textural stresses consisting of highly distorted and therefore thermally unstable atomic layers surrounding the fragments are created. The work hardening itself, that is, the changes in the structure of the polycrystalline aggregate and the resulting changes in deformational response and mechanical properties are therefore both time- and temperature-sensitive. The more extensive the crystal fragmentation and the larger the volume of distorted atomic layers created in the course of the deformation, the lower the thermal stability of the fragmented crystal pattern; the more pronounced therefore the effects of time and temperature.

The apparent continuity of most of the observed work-hardening diagrams is the result of the application of relatively high

strain rates, which prevent the manifestation and observation of discontinuities in the diagram. The lower the applied rates of deformation or loading, the more pronounced the discontinuous appearance of the work-hardening diagram (Fig. 46·1). The gradual refinement of the crystal grains by fragmentation will therefore tend to make the discontinuous shape of the diagram the less pronounced, the larger the deformation.

On the other hand, in the very early stages of deformation the slip planes are very finely distributed and crystal deformation goes on without disruption of the slowly deforming intercrystalline boundaries; during these stages the work-hardening diagram, therefore, will be the more perfectly continuous, the slower the loading rate. The existence of both effects has been observed in polycrystalline metals (Fig. 46·2).<sup>46·1</sup>

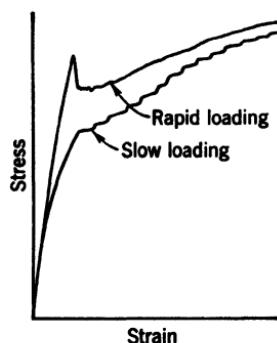


FIG. 46·1 Schematic stress-strain curves within the work-hardening range of a metal for rapid and slow loading.

existence of both effects has been observed in polycrystalline metals (Fig. 46·2).<sup>46·1</sup>

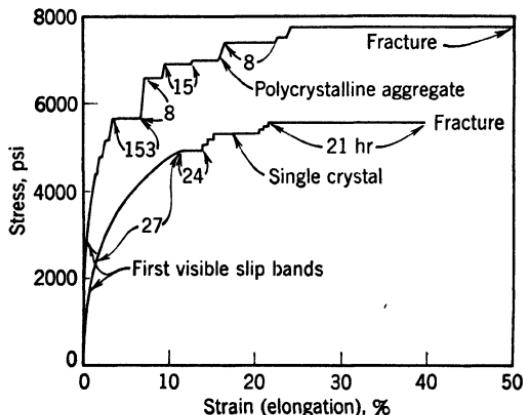


FIG. 46·2 Slow-loading work-hardening curves of aluminum at room temperature (after Hanson and Wheeler<sup>46·1</sup>).

If during the work-hardening process the *load* is controlled, the strain increments are discontinuous; under sustained load the deformation goes on at a decreasing rate until the newly mobilized

resistance of the fragmented pattern is able to carry the load (see Art. 20). If during the work-hardening process the *strain* rate is controlled, the resistance, that is, the stress increments, change discontinuously. Under a sustained constant strain, the stress necessarily drops, since at the moment of interruption of the deformation process the strain velocity suddenly becomes zero. Part of this drop, which is invariably observed in interrupted work-hardening tests (Fig. 46·3), can also be conceived as a relaxation of stress under sustained constant strain, by the adjustment of the relaxing grain boundaries to the imposed

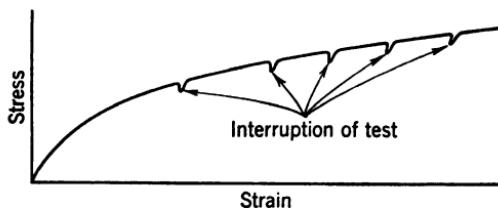


FIG. 46·3 Constant-strain-rate work-hardening curve of polycrystalline copper with interruptions of deformation (after Elam<sup>16·2</sup>).

strain. The drop in resistance immediately on interruption of the test is the smaller, the slower the strain rate applied, since it expresses the difference between the resistance to deformation at the applied strain rate and at zero strain rate. The additional drop that depends on the length of the period of interruption is, when it occurs, an expression of either relaxation or of the thermal instability of the fragmented crystal structure or of both effects. This drop will necessarily be the more pronounced, the nearer the temperature of the deformation process is to the recrystallization temperature.

If, after interruption at a certain strain, the deformation process is restarted at the previous strain rate, the stress at which further irrecoverable deformation sets in will be the stress at which the deformation process was stopped, only if the behavior of the metal is consistent with the concept of an equation of state. Rigorously this concept is inconsistent with the work-hardening process of most of the technically important polycrystalline metals. In those metals the stress at which plastic deformation starts at the previous strain rate after the interruption of the test is usually higher than the stress recorded at the

moment of interruption; the excess of this new yield limit over the resistance to plastic deformation at the moment of interruption depends on the extent of fragmentation, the applied strain rate, and the temperature; it increases with increasing duration

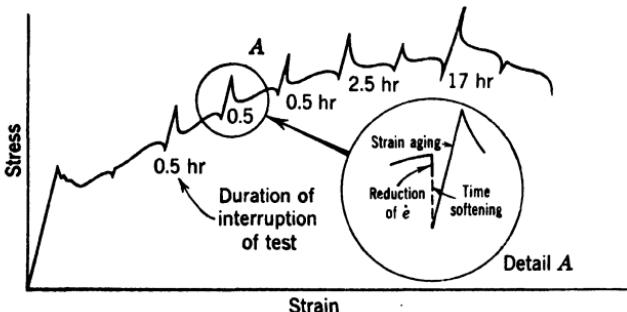


FIG. 46.4 Work-hardening diagram at constant strain for mild (aging) steel, with interruption of test (after Elam<sup>46.2</sup>).

of the period of interruption (Fig. 46.4). The more extensive the process of fragmentation at the time of the interruption, the higher the momentary increase in the yield limit. Thus, a more pronounced increase must be expected in initially coarse-grained

than in fine-grained aggregates (Fig. 46.5), whereas the increase will be less pronounced near the end of the deformation process than at its beginning, because of the largely completed grain refinement at the later stage.

The increased resistance, due to strain aging, however, is not permanent. With proceeding deformation it is usually followed by a drop back into a curve that is a continuation of the preceding

FIG. 46.5 Work-hardening diagram for nonaging steel (with rest periods.)

stress-strain diagram. The momentarily increased resistance to plastic deformation following an interruption of the deformation, as a result of which an excess load is needed to initiate further slip and fragmentation, has thus the character of an *upper yield limit* which is frequently observed in conventional tensile tests of steel; this load, however, is not needed to sustain the

progress of deformation, which is essentially a continuation of the interrupted fragmentation process.

Various phenomena may be responsible for the changes in the deformational response of the grain boundaries and the slip bands producing strain aging, such as precipitation and enforcing of atomic volume fit of the precipitated atoms, or formation of a different intercrystalline structure. The analysis of those effects forms the subject of various metallurgical theories of age hardening<sup>46-8</sup> the physical basis of which has been discussed in Art. 16.

The relative importance of the softening and the hardening effect will depend on the type of the metal, its content of alloying elements and initial structure, its melting point, the temperature during deformation as well as the temperature during rest periods, the extent of fragmentation, and the length of time since the interruption of the work-hardening process. Usually both effects are operative; they act, however, with different intensity at different times during the test. Whereas strain aging involves diffusion of foreign particles, softening is due to self-diffusion. Hence, the percentage of alloying or foreign elements appears to be the most important single factor responsible for strain aging.

The higher the temperature of the test and the more extensive the preceding fragmentation of crystals, the more pronounced the thermal instability of the fragmented crystal structure, and the higher therefore the rate of self-diffusion. At or near recrystallization temperature, the work-hardening effect is fully counterbalanced by the simultaneous resoftening, unless the rate of fragmentation is so high as to exceed the rate of recrystallization.

#### 47. Phenomenological Analysis of Plastic Deformation with Work Hardening

In the analysis of problems of plastic deformation with work hardening, a certain assumption must be made concerning the form of the relation connecting the second invariants of the deviators of stress and velocity strain, or of the stress and velocity strain intensities. This relation may be given by an empirical curve (Fig. 47·1) or the corresponding analytical function,

$$s_r = f(\dot{e}_r) \quad (47\cdot 1)$$

or, according to eq. 42·8,

$$(s_{ii} - p) = 2G'(\dot{e}_{ii} - \dot{e}_v) \quad \text{and} \quad s_{ij} = 2G'\dot{g}_{ij} \quad (47\cdot 2)$$

where  $G'$  denotes the shear modulus which, within the work hardening range, is a function of the stress intensity. For an incompressible material  $e_v = 0$ , and

$$(s_{ii} - p) = 2G' de_{ii} \quad \text{and} \quad s_{ij} = 2G' dg_{ij} \quad (47 \cdot 3)$$

For a state of uniaxial stress  $s_2 = s_3 = 0$ , and  $p = \frac{1}{3}s_1$ ; hence, if the strain increment is replaced by the strain,

$$s_1 = 3G'(s_1)(e_1 - e_v) = 3G'(s_1)e_1(1 + \mu) \quad (47 \cdot 4)$$

or, for volume-constant deformation,

$$s_1 = 3G'(s_1)e_1 \quad (47 \cdot 5)$$

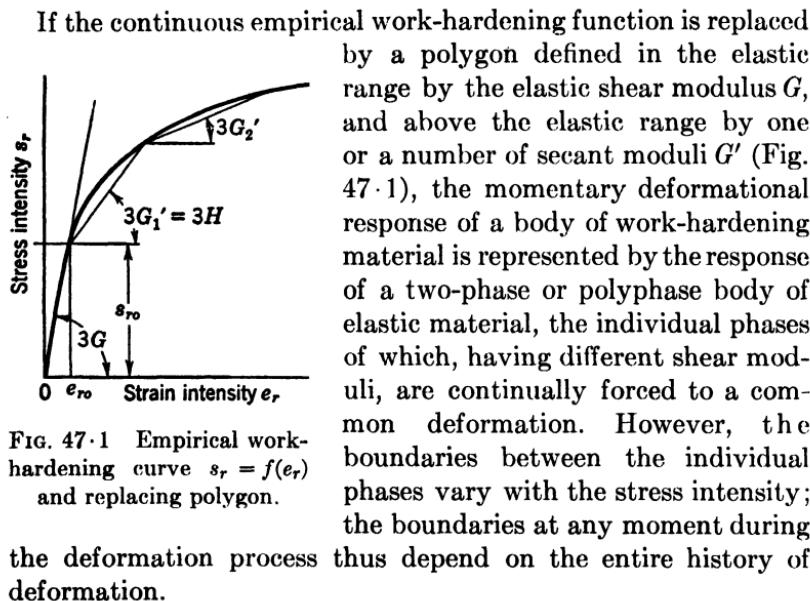


FIG. 47.1 Empirical work-hardening curve  $s_r = f(e_r)$  and replacing polygon.

the deformation process thus depend on the entire history of deformation.

In the simplest case of homogeneous volume-constant straining a "polygon" may be defined by the elastic shear modulus and one work-hardening shear modulus  $G' = H$ .<sup>47.1</sup> In this case eq. 47.1 may be written in the form,

$$(s_r - s_{0r}) = 3H(e_r - e_{0r}) \quad (47 \cdot 6)$$

where  $s_{0r}$  and  $e_{0r}$  denote the intensity of stress or strain, respectively, at the yield limit. For uniaxial stress  $s_r = s_1$ ,  $s_{0r} = s_{01}$ ,  $e_r = e_1$ ,  $e_{0r} = e_{01} = s_{01}/3G$ ; hence,

$$e_1 = \frac{1}{3H} (s_1 - s_{01}) + \frac{1}{3G} s_{01} \quad (47 \cdot 7)$$

For a compressible material, according to eq. 47·4 and because of  $e_r = \frac{2}{3}e_1(1 + \mu)$ , the total strain,

$$e_1 = \frac{1}{2H(1 + \mu)} (s_1 - s_{01}) + \frac{1}{2G(1 + \mu)} s_{01} \quad (47 \cdot 8)$$

The plastic strain component  $e_{1p} = e_1 - e_{1e}$ , where the elastic strain

$$e_{1e} = \frac{1}{2G(1 + \mu)} s_1 \quad (47 \cdot 9)$$

Hence,

$$e_{1p} = \left[ \frac{1}{2H(1 + \mu)} - \frac{1}{2G(1 + \mu)} \right] (s_1 - s_{01}) \quad (47 \cdot 10)$$

Under general conditions of homogeneous strain with the principal strains  $e_1 \neq 0$ ,  $e_2 \neq 0$ ,  $e_3 \neq 0$ , the resulting strains  $\bar{e}_1$ ,  $\bar{e}_2$ ,  $\bar{e}_3$  in the directions of the coordinate axes are

$$\bar{e}_{1e} = e_{1e} - \mu(e_{2e} + e_{3e}) \quad (47 \cdot 11)$$

and

$$\bar{e}_{1p} = e_{1p} - \nu(e_{2p} + e_{3p}) \quad (47 \cdot 12)$$

where  $\bar{e}_1 = \bar{e}_{1e} + \bar{e}_{1p}$ . Since  $e_{1p} = e_1 - e_{1e}$ ,

$$\bar{e}_1 = e_1 - \mu(e_2 + e_3) + (\mu - \nu)(e_{2e} + e_{3e}) \quad (47 \cdot 13)$$

This general relation which may be written in a similar form for  $\bar{e}_2$  and  $\bar{e}_3$  is simplified by the assumption  $\mu = \nu \neq 1/2$ . According to observations, this assumption is more reasonable than the assumption  $\mu \neq \nu = 1/2$  which is generally made and which implies that in the work-hardening range the material flows with constant volume. Actually, all observations of the deformational behavior of metals above the elastic range show that the specific density has a definite tendency to decrease with respect to the density in the elastic state as a result of the microscopically inhomogeneous conditions of strain created by crystal fragmentation. However, because of the very small density change  $\mu = \nu$  may be introduced, and eq. 47·13 written in the conventional form:

$$\bar{e}_1 = e_1 - \mu(e_2 + e_3) \quad (47 \cdot 14)$$

Introducing eq. 47·8 and similar expressions for  $e_2$  and  $e_3$ , eq. 47·14 gives the relation:

$$\begin{aligned}\bar{e}_1 = & \frac{1}{2H(1+\mu)} [s_1 - \mu(s_2 + s_3) - s_{01} + \mu(s_{02} + s_{03})] \\ & + \frac{1}{2G(1+\mu)} [s_{01} - \mu(s_{02} + s_{03})] \quad (47\cdot15)\end{aligned}$$

and similar expressions for  $\bar{e}_2$  and  $\bar{e}_3$ .

The principal stress components at the yield limits  $s_{01}$ ,  $s_{02}$ ,  $s_{03}$  must evidently satisfy the yield condition 41·5, while the stress components at any point within the work-hardening region must satisfy a similar condition in which, however, the right-hand side is replaced by  $2s_r^2$ , where  $s_r$  is obtained from eq. 47·6.

### References

- 44·1 M. F. SAYRE, *Trans. Am. Soc. Metals* **24** (1936) 932.
- 44·2 D. HANSON and M. A. WHEELER, *J. Inst. Metals* **45** (1931) 249.
- 44·3 J. F. NYE, *Proc. Roy. Soc. A* **198** (1949) 190; **200** (1949) 47.
- 44·4 W. A. WOOD, *Proc. Roy. Soc. A* **172** (1939) 231.
- 44·5 W. A. WOOD, *Proc. Phys. Soc.* **42** (1940) 39.
- 44·6 S. I. SMITH and W. A. WOOD, *Proc. Roy. Soc. A* **182** (1943–44) 404.
- 45·1 R. L. WOOLLEY, *Conf. Strength of Solids, Bristol 1947* *Phys. Soc. London* (1948) 51.
- 45·2 G. I. TAYLOR, *Proc. Roy. Soc. A* **145** (1934) 362, 388, 405.
- 45·3 A. M. FREUDENTHAL and M. REINER, *J. Applied Mechanics* **15** (1948) 265.
- 46·1 *Loc. cit.* 44·2.
- 46·2 C. F. ELAM, *Proc. Roy. Soc. A* **165** (1938) 579.
- 46·3 A. H. GEISLER, *Metals Technol.* (1948) T.P. 2436.
- 47·1 K. H. SWAINGER, *Phil. Mag.* (7) **36** (1945) 443.

C H A P T E R  
9

## CREEP AND RELAXATION

### 48. Creep of Viscoelastic Materials

Creep and relaxation are frequently defined as the deformational responses to constant stress and to constant strain, respectively, of a statistically isotropic material of unordered atomic or molecular structure. This definition, however, is too narrow to cover all aspects of the phenomenon for the designation of which the terms *creep* and *relaxation* are generally used.

In dealing with this general problem of the time dependence of deformation under sustained stress or constant load, and of stress or load under sustained initial deformation, the essentially amorphous viscoelastic substances are only one particular group of materials in which time effects are of considerable importance. The other group for which, under certain conditions, these effects may become very significant are the metals, both polycrystalline and deformed single crystals. The fact that the phenomenological behavior under sustained stress, load, or deformation of those two groups of materials of intrinsically different structure is designated by the same name because it shows a similar trend though being different from a structural aspect, has been responsible for a certain confusion in the evaluation and interpretation of creep and relaxation experiments and in the generalizations based thereupon. It has also introduced a certain complexity in the approach to a phenomenological theory of creep, since the necessity of defining the observed similarities and differences in the mechanical response of the internal structure to imposed constant load, stress or strain, and of isolating the few essential

factors governing the phenomenon has been frequently neglected in favor of procedures of devising empirical relations, reproducing results of series of rather uncoordinated experiments.

The complexity of the phenomena of creep and of relaxation for various materials and the difficulties in the interpretation of test results are in direct relation to the complexity of the internal structure of the considered material. It has been shown (see Art. 35) that the total strain of the Maxwell body under a uniaxial force consists of the elastic component  $e_0$  and the viscous (creep) component  $e_c$ , or

$$e = e_0 + e_c = \frac{s}{E} \left( 1 + \frac{t}{\tau} \right) \quad (48 \cdot 1)$$

where  $\tau = E/\lambda = 3G/3\eta$  denotes the *relaxation time*, which can

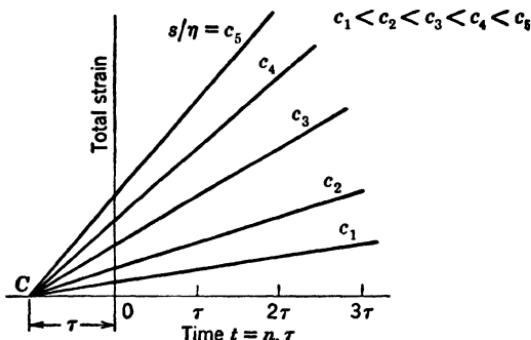


FIG. 48·1 Creep diagrams of a Maxwell body.

be defined either as the time after which the stress under a sustained constant strain has fallen off to  $(1/e)$  of its initial value, or as the time required to produce, under a sustained stress, an inelastic strain equal to the elastic strain. The family of creep curves  $e_c = f(t)$  of the Maxwell body thus consists of a family of straight lines passing through a common point  $C$  on the time axis to the left of the origin, the abscissa of which is equal to the relaxation time, or  $\overline{OC} = -\tau$  (Fig. 48·1).

The analysis in Art. 37 of the simple model consisting of two Maxwell elements coupled in parallel has shown that the linearity of the creep diagram of the simple viscoelastic body vanishes if the single relaxation time  $\tau$  of the Maxwell body is replaced by two or a number  $i$  of relaxation times  $\tau_i$  of Maxwell elements

coupled in parallel or of Kelvin elements coupled in series. Under constant stress, materials represented by such models creep at a gradually decreasing rate, approaching asymptotically a constant minimum rate defined by the longest relaxation time; thus, the creep diagram appears practically linear over a considerable range. The length of the initial range of creep at decreasing rate, which is generally called the primary or short-time creep, depends on the number of different relaxation times (or retardation times if Kelvin elements are coupled) and increases with their number. The model analysis has shown the behavior within this range to be essentially the result of a superposition of after-effects (Kelvin phases), each of which is governed by a function of the form  $(1 - e^{-t/\tau_i})$ . In terms of a continuous distribution  $F(\tau)$  of retardation times the part of the creep diagram produced by aftereffect may thus be described by a function of the form,

$$e_{c1} = e_0 \int_0^{\infty} F(\tau) (1 - e^{-t/\tau}) d\tau \quad (48 \cdot 2)$$

where  $e_0$  is the ordinate of the intersection of the asymptote of  $e_c$  with the strain axis, the so-called *zero intercept*. The total creep  $e_c$ , which is the sum of the primary creep or aftereffect  $e_{c1} = e_0 f(t)$ , and of the linear viscous creep  $e_{c2}$  (Fig. 48·2)

$$\begin{aligned} e_c &= e_0 f(t) + e_{c2} t \\ &= e_0 \left[ \int_0^{\infty} F(\tau) (1 - e^{-t/\tau}) d\tau + \frac{t}{\tau_1} \right] \end{aligned} \quad (48 \cdot 3)$$

where  $\tau_1$  denotes the relaxation time of the Maxwell phase.

The temperature dependence of the coefficient of viscosity  $\eta$  and thus of the retardation or relaxation times  $\tau = \eta/G$ , according to eq. 19·6, may be expressed by

$$\tau = \tau_0 e^{Q/RT} \quad (48 \cdot 4)$$

where  $\tau_0$  denotes a constant of dimension of time and  $Q$  the

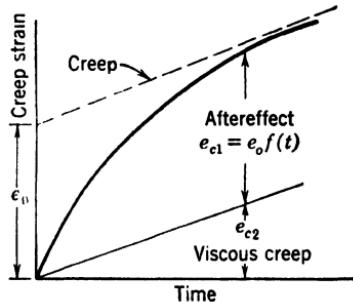


FIG. 48·2 Creep diagram produced by after-effect (recoverable) creep and viscous (irrecoverable) creep.

activation energy per mole; the creep strain is therefore strongly temperature-sensitive. The total strain increases in proportion to the decrease, with increasing temperature, of the coefficient of viscosity, and the shape of the primary creep function becomes rapidly very steep, approaching elastic behavior (Fig. 48·3).

A large number of viscoelastic materials, such as, for instance, the photoelastic materials and many textile fibers, do not show real creep, only delayed recoverable deformation. In this case the observed creep rate approaches zero at infinite time. If in

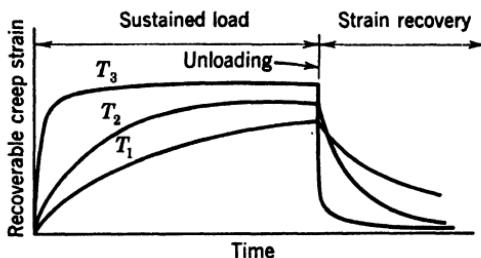


FIG. 48·3 Creep and recovery curves for fully recovering viscoelastic material as a function of temperature ( $T_3 > T_2 > T_1$ ).

the course of the deformation process no irrecoverable changes take place in the molecular structure of the material, such behavior is due only to the existence within the material of a continuous elastic network which cannot be permanently deformed (see Art. 37). The same phenomenon of decreasing creep rate may be due, however, to a different cause when the real creep is accompanied by irrecoverable changes in the molecular structure resulting from chemical reactions which accompany the deformation. Such changes may be the formation of additional cross links between the molecular chains, or the breaking and re-formation of links, or the volatilization of part of a constituent fluid phase, or crystallization (see Art. 20). All these changes by increasing in the course of time the apparent rigidity of the material produce a gradual reduction of the creep rate which will necessarily be the more significant, the longer the duration of the loading.

The difference between the purely mechanical after-effect and the effects of permanent change of molecular structure in producing a decreasing creep rate is not observable in the course of

loading. It becomes visible only during unloading, since the inelastic deformation due to the mechanical after-effect is fully recoverable after sufficiently long (theoretically infinite) time, whereas genuine creep, the rate of which is affected by gradual changes within the molecular structure, is unrecoverable. The shape of the unloading diagram is therefore an indication of the relative importance, in viscoelastic materials, of the primary creep (that is, the after-effect) and of the secondary, not necessarily linear stage of genuine creep (Fig. 48·4). The after-effect is frequently referred to as *creep recovery*.

On the other hand, changes in the molecular structure accompanying deformation may also produce a more or less gradual increase of the creep rate, if such changes consist in the rupture of cross links during deformation or in a local reduction of the extent of polymerization (depolymerization). As a result of the

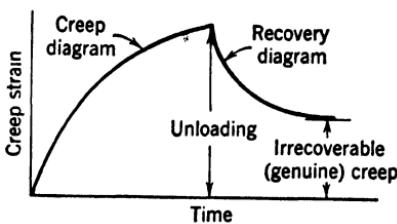


FIG. 48·4 Irrecoverable creep determined from recovery diagram.

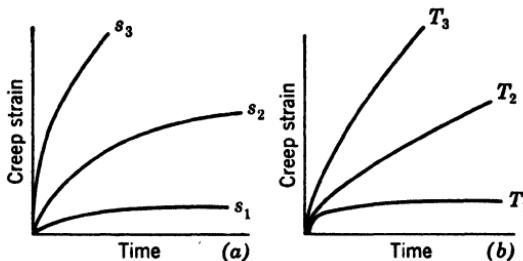


FIG. 48·5 Variation of creep curves of viscoelastic material (a) with stress ( $s_3 > s_2 > s_1$ ) and (b) with temperature ( $T_3 > T_2 > T_1$ ).

reduced rigidity of the network of chain molecules, an increased rate of deformation within the viscous phases is required to balance the applied stress. Since the extent of structural changes during deformation depends on stress, temperature, and time, different shapes of creep curves are obtained for the same viscoelastic material if any one of the above parameters is varied (Fig. 48·5).

Although the term creep implies, in general, deformation at sustained constant stress, actual creep tests are usually performed under constant load, the recorded creep diagrams being in fact deformation-time diagrams at constant load but varying stress.

In the tension test of a specimen of circular cross section  $A_0$  of a volume-constant Maxwell body under constant load  $P$ , the strain velocity of the moving end of the specimen:

$$\dot{e}_1 = \frac{d}{dt} \left( \frac{dl}{l} \right) = - \frac{d}{dt} \frac{dA}{A} = - \frac{1}{A} \frac{dA}{dt} \quad (48 \cdot 5)$$

By introducing eq. 48·5 and the stress  $s = P/A$  into the equation of the Maxwell body 35·29, the relation is obtained:

$$\frac{1}{A} \frac{dA}{dt} = - \frac{1}{E} \frac{d}{dt} \left( \frac{P}{A} \right) - \frac{1}{\lambda} \left( \frac{P}{A} \right) \quad (48 \cdot 6)$$

Integration of eq. 48·6 for  $P = P_0$  under the boundary condition  $A = A_0$  for  $t = 0$  leads to the equation,

$$\frac{A}{A_0} - e_0 \ln \left( \frac{A}{A_0} \right) = 1 - e_0 \frac{t}{\lambda} \quad (48 \cdot 7)$$

with  $e_0 = P_0/EA_0$ . For not too large strains eq. 48·7 may in first approximation be replaced by

$$\frac{A}{A_0} = 1 - \dot{e}_0(t/\lambda) \quad (48 \cdot 8)$$

or, because of constant volume,

$$\frac{l}{l_0} = \frac{1}{1 - \dot{e}_0(t/\lambda)} \quad (48 \cdot 9)$$

Hence, the logarithmic strain in the constant-load tensile test,

$$e = \ln \left( \frac{l}{l_0} \right) = \ln \left( \frac{1}{1 - \dot{e}_0(t/\lambda)} \right) \quad (48 \cdot 10)$$

In the constant load compression test

$$e = \ln \frac{1}{1 + \dot{e}_0(t/\lambda)} \quad (48 \cdot 11)$$

For constant stress  $P/A = \text{const}$ , eq. 48·6 becomes

$$-\frac{dA}{A} = \frac{1}{\lambda} \left( \frac{P}{A} \right) dt \quad (48 \cdot 12)$$

and, integrated,

$$e = \ln \left( \frac{A_0}{A} \right) = \ln \left( \frac{l}{l_0} \right) = e_0 \left( \frac{t}{\lambda} \right) \quad (48 \cdot 13)$$

In Fig. 48·6 eqs. 48·10, 48·11, and 48·13 are compared for

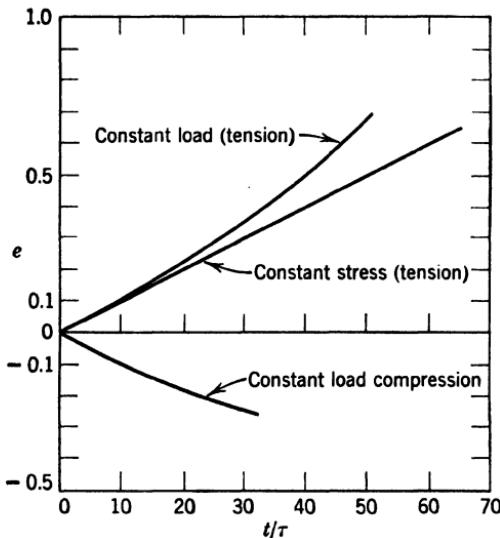


FIG. 48·6 Comparison of creep curves of viscoelastic bar in constant-load tension test, constant-load compression test, and constant-stress test.

$e_0 = 0.033$ . With increasing time the difference between the time-strain curves becomes considerable. The results of constant-load-creep tests must therefore not be confused with those of constant-stress-creep tests, unless the total creep-strain is small. The observed creep rates are thus applicable only to the specific testing conditions.

In attempting to apply the results of uniaxial creep tests to general states of stress the fact should not be ignored that the rate of creep in a certain direction is in general not related to the axial stress component in the same direction, but to the entire deviator of stress, since viscous flow is a response to distortional

stresses only. Hence, the smaller the share of the deviatoric stresses in the general state of stress, the smaller the rate of creep. Therefore the creep rate in the directions of the principal stress  $s_1$  is reduced by stresses  $s_2$  and  $s_3$  of the same sign as  $s_1$ , becoming zero for  $s_1 = s_2 = s_3$ , but is increased by stresses  $s_2$  and  $s_3$  of opposite sign.

#### 49. Creep of Metals

The complexity of creep phenomena in polyphase materials, such as polymers and polycrystalline metals, is due to the interaction of the constituent phases during the deformation, and to the difference in the responses of those phases to changes of parameters of the test, that is, to changes of stress, time, and temperature. However, although even in highly complex viscoelastic materials, such as high polymers, the interaction between an elastic network and a number of viscous phases produces phenomena that are essentially combinations of after-effects and of viscous creep, the creep of metals is generally a combined effect of the predominantly viscous inelastic deformation within the unordered intercrystalline boundaries and the complex deformation by slip and fragmentation of the ordered crystalline domains. The phenomenon is made still more complex by the change of structure within the crystalline regions produced in the course of the deformation and by the thermal instability of the deformed structure (see Art. 20).

The character of the interaction between the crystalline and the intercrystalline phases depends on the relative rigidities of the intercrystalline boundaries and the crystalline domains. As long as the inelastic deformation is concentrated within the crystal boundaries which restrain the crystalline domains from deforming by slip, the metal responds essentially like a relatively simple viscoelastic material, in which the response of the crystals is elastic. The extent of creep under such conditions is limited and is governed by viscoelastic relations of the type of eq. 48·3. Such conditions are present in metals subject to moderate stresses at all temperatures, at which the slip resistance of the crystals exceeds the resolved shear stress; they may therefore exist at elevated temperatures, if either the applied stress is so low or the slip resistance within the crystal region so high that, in spite of the sharply reduced rigidity of the intercrystalline regions and

the consequent rather rapid transfer of the response to the external load into the crystal domains, no slip is produced. The material then responds essentially as a viscoelastic body which deforms by creep of the boundary material and by relative motion and rotation of the grains along the boundaries.

Conclusive evidence of such rotation has been obtained by Moore and coworkers in creep tests of lead.<sup>49-1</sup> When polished specimens were scratched with parallel lines and then drawn several percent at room temperature the scratches assumed different orientation in adjacent grains when the extension proceeded at a slow rate but maintained the same orientation when the extension was rapid. Thus, *disorientation* of crystalline domains due to their rotation within the viscously yielding disordered regions (grain boundaries, slip bands) appears to be a phenomenon which is associated with the deformation of polycrystalline metals under conditions in which those regions appear relatively soft.<sup>49-2</sup> To produce such conditions loads must not necessarily be of long duration if the order of magnitude of relaxation times of the intercrystalline regions is also relatively short.

The essentially viscoelastic character of the creep is changed when either the slip resistance of the crystal regions is so far reduced or the resolved shear stress so far increased that the rotation of the crystals within the boundary material is followed by slip and fragmentation of a certain number of crystals. The character of the creep thus changes gradually from a purely viscoelastic deformation, to a highly complex combination of viscous disorientation, slip, work hardening, recovery or recrystallization, and progressive local fracture.

Viscous creep of the polycrystalline metal aggregate associated with rotation of grains along the grain boundaries is thus characteristic for the initial stage of creep tests performed at a stress, which is not high enough to produce slip within the crystal regions by local disruption of intercrystalline boundaries immediately on application of the load. The length and importance of this stage evidently depends on the specific volume of grain-boundary material and on the resilience of the individual grains. This type of creep is the more important, the smaller the grain size of the metal, the larger the specific volume of the boundary regions, and the higher the resilience of the individual grains.

Since under the same stress level the extent of crystal slip and fragmentation in fine-grained metals is much smaller than in coarse-grained metals, while the volume of viscous material is larger, the viscoelastic stage of creep is longer and of greater importance in fine-grained than in coarse-grained polycrystalline metals. At sufficiently low stresses the viscous creep component in coarse-grained metals may gradually vanish and creep thus proceed at a decreasing rate that gradually approaches zero as further grain-boundary deformation is blocked by contact and interlocking of grains. Generally, therefore, the rate of steady creep is higher in fine-grained than in coarse-grained metal, at least at the moderate stresses at which crystal slip is limited or practically nonexistent. Hence, no creep at stresses below the critical shear stress should be expected in bodies made up of single crystals. It is for this reason that, for instance, tungsten filaments, exposed to very high temperatures in lamps, are now made of single crystals, since their efficiency depends on the total absence of creep (sagging).

At a stress above the critical shear stress the lattice of the single crystal is distorted as a result of slip, and thus a viscous component is introduced into the subsequent deformation process which becomes the more significant the more extensive the slip. Thus, creep of a single crystal is initiated by slip and is therefore, in the first stage, as time-sensitive as the slip process itself. Hence, the precipitation of foreign particles within the slip planes retarding and finally blocking slip, as well as the diffusion of those particles under the applied stress by which the slip is unblocked after a certain delay which depends on the diffusion rate (see Art. 20) will necessarily modify not only the creep rate of the single crystal, but also that of the aggregate made up of such crystals.

Since the viscous component of creep of the single crystal is introduced as a result of the disordered atomic layers around the slip planes, viscous creep of crystals can not occur without previous plastic deformation, and it can only be maintained by continued slip; otherwise it is rapidly blocked as a result either of selfdiffusion by which the disorder around the slip planes is gradually eliminated, or of precipitation as a result of which the slip resistance is raised on the slip planes. There is therefore a close interrelation between the plastic slip and the viscous

component in single crystals as well as in polycrystalline aggregates.

When creep of the polycrystalline aggregate is due to the combined effect of viscous deformation of the grain boundaries and distorted atomic layers, of rotation of crystals and of slip, the decreasing rate at which creep proceeds during the first or transient stage is no longer the result of pure after-effect, that is, of the internal redistribution of the mechanical response to the load between the elastic and the viscous phase of a material of unchanging structure; it is the expression of the combination of work-hardening fragmentation, of viscoelastic redistribution of the deformational response to the load within the changing crystal structure, and of precipitation. Moreover, with increasing deformation the tendency to viscous creep is increased because of the increasing volume of disordered material produced by the fragmentation and because of the intensified thermal instability of the fragmented structure. This instability promotes local recovery and recrystallization, accompanied by accelerated deformation; since the rate of recrystallization depends on the extent of work hardening, a periodic fluctuation of acceleration and deceleration of creep must be expected quite independent from effects of precipitation. Such fluctuations express the alternations of work-hardening and recrystallization periods, as the recrystallization rate, which exceeds the work-hardening rate at a certain strain, sharply drops after partial recrystallization has taken place by which the work-hardening rate is again raised. The existence of such periodicity in the deformation process could actually be deduced from X-ray diffraction studies.<sup>49-3</sup>

Because of the pronounced influence of temperature on both the viscosity of the distorted atomic layers and the rate of recovery and recrystallization within the fragmented structure, the rate of the viscous creep component increases very sharply with temperature. It also increases with the stress level, because of the increased extent of fragmentation produced by higher stress. Thus the shapes of the creep curves will vary with the applied stress level and with temperature, as shown schematically in Fig. 49·1; the higher the temperature and the stress level, the more nearly viscous the character of the creep, whereas low temperatures and low stresses produce a predominance of the work hardening and of the after-effect.

As in the case of viscoelastic materials, the respective influences of structural changes and of genuine after-effect on the decrease of the creep rate cannot be separated during the loading period, but only by observing the strain recovery on load removal. As

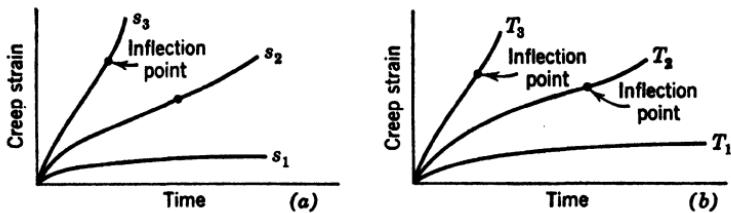


FIG. 49·1 Variation of creep curves of metals (a) with stress ( $s_3 > s_2 > s_1$ ), and (b) with temperature ( $T_3 > T_2 > T_1$ ).

creep, which is the result of permanent structural change by work hardening and recrystallization, cannot be recovered, the influence of the after-effect is represented by the recoverable part of the inelastic deformation produced by the load.

So far only two ranges of creep have been considered: the *primary* or transient creep at decreasing rate, and the *secondary* or viscous creep at practically constant rate. Creep tests at moderate and high stresses invariably show that a third final range exists, during which the creep rate increases steadily, and which sooner or later is terminated by fracture. This range has been indicated in Fig. 49·1 which shows that, in general, creep

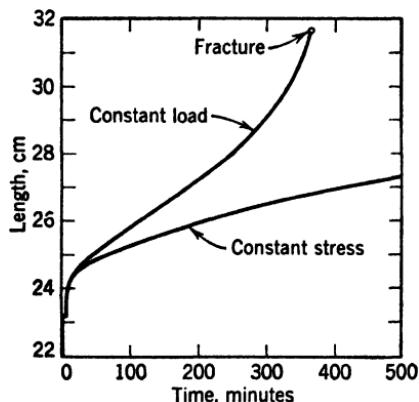


FIG. 49·2 Creep curves of lead wire under constant load and under constant stress (after Andrade<sup>49·4</sup>).

curves of metals have a definite inflection point, which delimits the "third stage."

The third stage of creep was believed for a time to be an exclusive characteristic of the constant-load test, which could be eliminated by keeping the stress constant. Such belief was sup-

ported by the results of comparative creep tests on metal wires performed by Andrade and reproduced in Fig. 49·2.<sup>49·4</sup> It has become increasingly evident, however, that the accelerated creep of the third stage is due not only to the increasing stress over the contracting area in the constant load test, but also to real changes within the structure of the material. These changes may be the effect of intensified recrystallization resulting from extensive work-hardening fragmentation by which the subsequent rate of creep by further slip and fragmentation is increased, or they may result from the fact that grain-boundary deformation and rotation of neighboring crystal grains without slip cannot go on indefinitely without the opening of a large number of small cracks within these boundaries. Hence, in the latter case third-stage creep does not represent a pure deformation process, but a process of progressive damage (Fig. 49·3). However, also in the case of third-stage creep produced by intensified relaxation the sharp decrease in the work-harden-ing rate will rapidly terminate the process of homogeneous deformation by causing necking (see Art. 37) and thus initiating a range of instable deformation leading rapidly to fracture. Thus third-stage creep, both by grain-boundary deformation and by slip, can be considered a stage of incipient or progressive fracture, fundamentally different from the two preceding stages of deformation.

Since the length of the second stage of creep depends on the

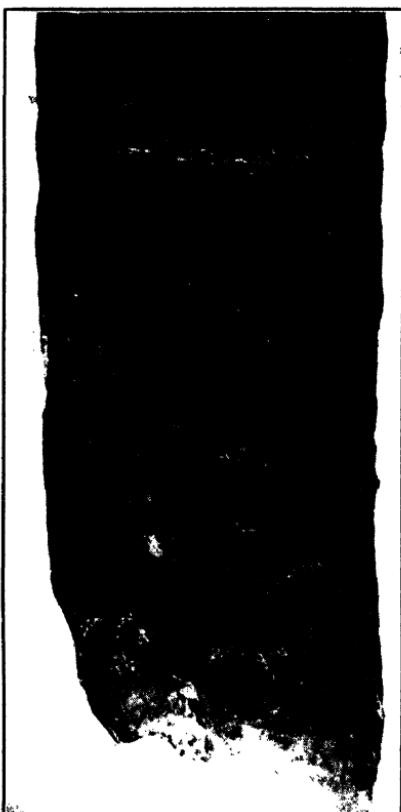


FIG. 49·3 Cracking characteristic of third-stage creep in lead (after Moore and Dollins;<sup>49·5</sup> courtesy Prof. C. W. Dollins).

rate at which creep proceeds during this stage, it will be the shorter, the higher this rate. Therefore third stage creep in a constant load test will set in much earlier than in a creep test in which the initial stress or the strain rate is kept constant. Because of the additional effect of stress increase by area reduction, the rate of third-stage creep in tension increases much more rapidly in the constant-load test than in the constant-stress test. There is, however, no reason to assume that the genuine third stage creep can be eliminated by keeping the stress constant; it can only be considerably retarded.

The length of the third-stage creep is an indication of the character of fracture; the more pronounced the increase in creep rate, the larger the creep at fracture, and the shorter the duration of this stage, the larger is the extent of crystal slip and fragmentation preceding fracture; the more ductile, therefore, the rather rapidly developing fracture. A long and very slowly accelerated third stage, on the other hand, indicates a slowly developing intercrystalline cracking followed by sudden and brittle fracture concentrated within the grain boundaries. Frequently both types of fracture can be observed to develop simultaneously.

Since the third stage of creep represents a condition of progressive or incipient fracture, the approach, under service conditions, of the creep strain to the inflection point of the creep diagram which defines the insetting critical third stage condition must be prevented with reasonable safety. The analysis of the third stage of the creep function, therefore, does not form part of the analysis of deformational behavior, but belongs to the analysis of progressive fracture. As far as design for creep is concerned, this stage is beyond the critical limit of deformation. Creep design should therefore be based on the two-stage creep curve limited by the inflection point.

Under conditions for which the creep of a polycrystalline metal is due to the combined effect of the viscosity of the grain boundaries and of the structural changes produced by work hardening, the linear relation between creep rate and uniaxial stress  $\dot{\epsilon}_c = s/\lambda$  must be replaced by a relation more suited to reproduce the combined effects within the intercrystalline and the crystalline regions. Such a relation has been derived in Art. 19 on the basis of structural considerations; it can be written in the simplified form,

$$\dot{e}_c = c \sinh \left( \frac{s}{s_0} \right) \quad (49 \cdot 1)$$

where  $c$  and  $s_0$  are experimental constants of the dimension of strain velocity and stress, respectively. Equation 48·3 can therefore be written in the form:

$$e_c = e_0 f(t) + c \sinh (s/s_0) t \quad (49 \cdot 2)$$

The extent of the variation of the *zero intercept* with stress has not been reliably established by tests. However, for moderate stresses it might be assumed, as a rough approximation, that the asymptotically reached second stage represents a state of predominantly viscous creep with a constant stress-independent coefficient of viscosity. In this case the asymptotes to the creep curves would pass through a single point on the  $t$  axis, left of the origin; with  $f(t) = 1$  for  $t = \infty$  in eq. 49·2, the zero intercept  $e_0$  could thus be derived from the relation  $e_0 = \text{const } \dot{e}_c$ . Both the steady creep rate and the zero intercept would thus be represented by the same hyperbolic sine relation, and the general creep function could be written in the form:

$$e_c = \left[ c_1 \int_0^{\infty} F(\tau) (1 - e^{-t/\tau}) d\tau + c_2 t \right] \sinh (s/s_0) \quad (49 \cdot 3)$$

Most of the test results<sup>49·6</sup> suggest that in general the asymptotes of the creep curves do not intersect in a common point on the time axis; two hyperbolic sine functions, differing in the constant  $s_0$ , are therefore required to express the creep function in the most general form:

$$e_c = c_1 f(t) \sinh (s/s_{01}) + c_2 t \sinh (s/s_{02}) \quad (49 \cdot 4)$$

## 50. Analysis of Creep Problems

Analysis of creep problems is based on the mathematical representation of two-stage creep curves by expressions of the form 48·3 or 49·4 or simpler approximations, or by an approximation to the general creep law (eq. 49·4), neglecting the transient (primary) creep stage and using only the minimum rate of steady creep.

The simplest and most widely used curve-fitting expressions are power laws which may be either of the form,

$$\dot{\epsilon}_c = c_1 \left( \frac{s}{s_0} \right)^m \quad (50 \cdot 1)$$

where  $m > 1$ ,

$$s_0 = \left( \frac{\dot{\epsilon}_c}{c_2} \right)^n \quad (50 \cdot 2)$$

where  $0 < n < 1$ . These power laws reproduce creep behavior over a limited range only. For large stresses  $s$ , the expression,

$$\dot{\epsilon}_c = ce^{(s/s_0)} \quad (50 \cdot 3)$$

or its equivalent,

$$s = s_0 \log \left( \frac{\dot{\epsilon}_c}{c} \right) \quad (50 \cdot 4)$$

are introduced as approximations to the hyperbolic sine law; for small stresses  $s$  the linear relation of purely viscous creep  $\dot{\epsilon}_c = cs$  may be introduced as a fair approximation.

For general conditions of stress the creep functions may be written in terms of the *intensity of stress* and the *intensity of the creep rate* or in terms of the octahedral shear stress and rate of strain (see Art. 41). In the form of a power law, using the intensities,

$$s_r = c_1 (\dot{\epsilon}_r)^n \quad (50 \cdot 5)$$

This law includes both the linear viscous substance with  $n = 1$  and the perfectly plastic substance with  $n = 0$ .

In the form of the logarithmic law,

$$s_r = s_{r0} \log \left( \frac{\dot{\epsilon}_r}{c} \right) \quad (50 \cdot 6)$$

If eqs. 50·1 to 50·6 are used for the analysis of creep problems, the effect of the primary stage of creep is tacitly neglected; the solutions obtained represent conditions of steady-state creep. The approach has been chosen in practically all cases in which a numerical analysis of creep problems has been attempted. In a few instances only have attempts been made to consider the primary stage of creep by expressing the creep rate as the product of a power function of time and of a function of stress:

$$\dot{\epsilon}_c = \text{const } f(s)t^m \quad (50 \cdot 7)$$

Although the use of formulas of this type introduces considerable mathematical difficulties they must be applied in cases of high temperature design for short life when the steady-state creep is no longer of principal practical importance.<sup>50-1</sup>

The most serious difficulty in creep design is in deriving the expected rate of steady creep under the actual conditions from existing test results. For this purpose it is necessary either to design the test so as to duplicate conditions of service, or to extrapolate from the results of relatively short-time tests, or to devise a rational method of accelerated creep tests, using the fact that, within the range of small deformations, the deformational behavior of both viscoelastic materials and metals is governed by a combined strain-rate-temperature parameter (see Art. 19). In the last case a testing temperature is selected at which the deformation during the expected service life and at the service temperature can be reproduced within a testing period of reasonable length. Evidently the testing temperature must still remain within the range of validity of an equation of state so that no appreciable change of internal structure of the material is produced during the test.

The principal object of the accelerated test is the determination of the limiting strain which defines the inflection point of the creep curve between the second and the third stage. The limiting strain appears to be nearly constant (Fig. 50-1) as long as the range of temperature variation is below recrystallization temperature. It is the limiting strain and not the steady creep rate alone, on which a safe creep design depends; only within the range of steady creep, delimited by the critical strain is an extrapolation permissible from the results of tests extending over a few hundred hours to service times of several thousand hours.

In solving general creep problems the consideration of the linear elastic-viscoelastic analogy for incompressible materials, discussed in Art. 35, is of particular importance. This analogy may be extended to compressible elastic solids in all cases, in

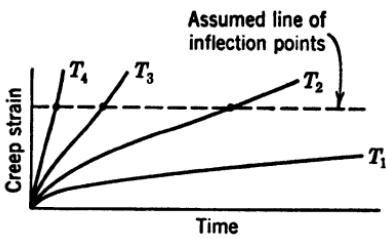


FIG. 50-1 Assumption for accelerated creep test. Temperatures ( $T_4 > T_3 > T_2 > T_1$ ).

which the stresses are independent of Poisson's ratio, as for instance in plane bending.

A similarity might thus also be assumed to exist between the equilibrium stresses in the loading stage in materials defined by a nonlinear stress-strain relation and the stresses under conditions of stationary flow in materials whose behavior is described by a similar nonlinear relation between stress and strain velocity.<sup>50-2</sup> Hence, the state of stress associated with conditions of creep, expressed by eq. 50·5 would be identical with the state of stress within a work-hardening material, the stress-strain relation of which is expressed by the power law:

$$\sigma_r = c_1 e_r^n \quad (50\cdot 8)$$

## 51. Relaxation

As long as creep is essentially viscoelastic and proceeds by homogeneous deformation within an essentially unchanging structure, a definite correlation exists between creep and relaxation of the same material and can theoretically be used in deriving relaxation curves from creep curves. It has been shown in Art. 38 that the rigorous evaluation of this relation may be difficult and may require the application of methods of operational analysis. However, if part of the observed creep is due to irrecoverable changes within the crystalline structure of metals or to identifiable changes within the molecular structure of viscoelastic materials, this relation no longer exists; the instantaneous values of the creep rate have become functions of strain and of previous strain history and thus of the amount of energy expended in the test.

Most of the difficulties encountered in the attempts to correlate creep and relaxation tests of real materials are probably due to the fact that the energy aspect of both tests is different; the rate of application of external energy in the creep test  $dA/dt = T_{00}\dot{E}_0 > 0$ , since  $\dot{E}_0 > 0$ , whereas in the relaxation test  $dA/dt = T_0\dot{E}_0 = 0$ , since  $\dot{E}_0 = 0$  because of  $E_0 = \text{const}$ . Only if none of the energy applied in the course of the creep test is expended in producing identifiable and irrecoverable changes of the internal structure of the material, does this difference in the energy aspect of the tests not affect the relation between creep and relaxation.

Relaxation phenomena govern the behavior of bolted assemblies under high-temperature service conditions. Although

special relaxation testing equipment has been built, relaxation tests are less easy to perform and therefore less common than the simple creep tests; it is therefore of practical importance to be able to derive the time-stress curve of a relaxing bolt from the data obtained by a creep test.

Under assumptions of simple relaxation, that is, for a constant length  $l$  of the bolt, gripping infinitely rigid flanges, the relation holds

$$\frac{Pl}{AE} + e_c l = s/E + e_i l = \frac{P_i}{AE} = s_i/E \quad (51 \cdot 1)$$

where  $P$  and  $P_i$  denote the momentary and the initial force in the bolt, respectively. In a simplified form eq. 51·1 may be written

$$e + e_c = e_0 \quad (51 \cdot 2)$$

where  $e$  and  $e_0$  denote, respectively, the elastic and the initial strain. Differentiating eq. 51·2, the relations between the strain rates are

$$\dot{e} + \dot{e}_c = 0 \quad (51 \cdot 3)$$

or

$$\dot{e}_c = -\dot{e} = -\frac{1}{E} \frac{ds}{dt} \quad (51 \cdot 4)$$

This is the differential equation of the transformation of creep and relaxation for ideal homogeneous isotropic materials.

Equation 51·4 is valid if the structure of the material is not changed during the creep test so that the momentary creep-rate depends on the momentary values of stress only, not on the value of strain and on strain history. With the steady-state creep function,

$$\dot{e}_c = f(s) \quad (51 \cdot 5)$$

integration of eq. 51·4 gives

$$t = \frac{1}{E} \int_s^{s_i} \frac{ds}{f(s)} \quad (51 \cdot 6)$$

Introducing the hyperbolic-sine law 49·1 for steady-state creep into eq. 51·6 gives

$$t = \frac{1}{Ec} \int_s^{s_i} \frac{ds}{\sinh(s/s_0)} \quad (51 \cdot 7)$$

The integration of this equation leads to a similar relation to that obtained as a solution of eq. 39·4:

$$\tanh(s/2s_0) = \tanh(s_i/2s_0)e^{-t/t_0} \quad (51\cdot8)$$

where  $t_0 = s_0/Ec$ ; hence,

$$t = t_0 \log \frac{\tanh\left(\frac{s_i}{2s_0}\right)}{\tanh\left(\frac{s}{2s_0}\right)} \quad (51\cdot9)$$

For linear viscous behavior with  $e_c = cs$ ,

$$t = \frac{1}{Ec} \log\left(\frac{s_i}{s}\right) \quad (51\cdot10)$$

If the power function 50·1 is used instead of the hyperbolic sine law, eq. 51·6 becomes

$$t = \frac{1}{Ec} \int_s^{s_i} (s/s_0)^m ds = \frac{s_0^m}{(m-1)cEs^{m-1}} \left[ 1 - \left(\frac{s}{s_i}\right)^{m-1} \right] \quad (51\cdot11)$$

From either of eqs. 51·9 or 51·10 a curve of stress  $s$  as a fraction of the initial stress  $s_i$  versus the time  $t$  required to produce the stress relaxation from  $s_i$  to  $s$  (relaxation curve) is easily obtained.

The relaxation problem of bolted assemblies becomes more complex if the elasticity of the assembly is introduced; the length of the bolt is then no longer a constant but becomes a function of the stress. The basic condition in this type of relaxation is that at any instant the momentary length of the bolt equals the length of the compressed assembly. Considering the assembly to undergo both elastic and inelastic deformation, the sum of the inelastic strains for bolt and assembly must, at any time, be equal to the sum of the initial elastic strains minus the sum of the momentary elastic strains. Thus eq. 51·2 applies, if the strains  $e$ ,  $e_c$ ,  $e_0$  are so defined as to refer to both bolt and assembly, considering their different creep curves and different elastic moduli.

In applying creep-relaxation transformation formulas it must be remembered that the question of the validity of eq. 51·4 always arises, since thermal stability of the mechanical structure of metals at elevated temperatures is the exception rather than

the rule. Microscopically defined structural stability itself is not a sufficient criterion of such validity, since recovery and precipitation become visible under the microscope only after they have reached a certain intensity, whereas they affect mechanical behavior already at a stage at which structural changes are not yet optically discernible.

### References

- 49·1 H. F. MOORE, B. B. BETTY, and C. W. DOLLINS, *Univ. Illinois Eng. Expt. Sta. Bull.* 272 (1935).
- 49·2 W. A. WOOD and W. A. RACHINGER, *J. Inst. of Metals* 76 (1949) 237.
- 49·3 W. A. WOOD, *Proc. Roy. Soc. A* 172 (1939) 231.
- 49·4 N. DA C. ANDRADE, *Proc. Roy. Soc. A* 90 (1914) 329.
- 49·5 H. F. MOORE and C. W. DOLLINS, *Univ. Illinois Eng. Expt. Sta. Bull.* 347 (1943) 25.
- 49·6 E. P. POPOV, *J. Applied Mechanics* 14 (1947) A-135.
- 50·1 L. F. COFFIN, P. R. SHEPLER and G. S. CHERNIAK, *J. Applied Mechanics* 16 (1949) 229.
- 50·2 A. NADAI, *v. Kármán Anniversary Volume* (1941) 237.

C H A P T E R  
**10**

## INELASTIC BEHAVIOR UNDER DYNAMIC CONDITIONS

### 52. The Damping Capacity

In the test or performance of materials under conditions of cyclic loading energy is dissipated during each load cycle in the interchange between potential and kinetic energy. This dissipation is manifest in various ways, for instance in the appearance of a *hysteresis loop* in the force-deformation or the stress-

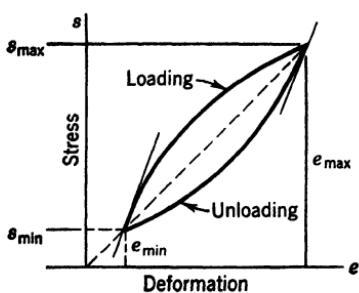


FIG. 52·1 Hysteresis loop under imposed periodic stress, fluctuating between  $s_{\max}$  and  $s_{\min}$ .

strain diagram under an imposed period force (Fig. 52·1), in the energy input required to keep the amplitude of a forced vibration constant, in the temperature increase of the material specimen during vibration, in the damping of the resonance curve; or in the decay with time of the amplitude of free vibrations of a specimen. The *specific damping capacity* is usually defined either in terms of forced

vibrations, as the ratio  $\psi$  between the energy per unit volume  $dW_D$  dissipated during a completely reversed cycle and the maximum potential energy  $W$  stored up in this volume during the cycle or, in terms of free vibrations, as the decrease, per cycle, of the natural logarithm of the amplitude  $a$  or of the energy  $W$

of the decaying vibrations, which is termed the *logarithmic decrement*  $\delta$ .

According to the foregoing definition,

$$\psi = \frac{dW_D}{\Delta N} \cdot \frac{1}{W} = \frac{d \log W}{\Delta N} \quad (52 \cdot 1)$$

where  $N$  denotes the sequential number of the load cycles and  $\Delta N = 1$ . If it is assumed that the energy is roughly proportional to the square of the amplitude  $a$ , or  $W = ca^2$ , eq. 52·1 may be written

$$\psi = \frac{d(ca)^2}{ca^2 \Delta N} = 2 \frac{d \log a}{\Delta N} = 2\delta \quad (52 \cdot 2)$$

The damping capacity of a material can also be determined by the observation of a resonance curve, which gives the amplitude

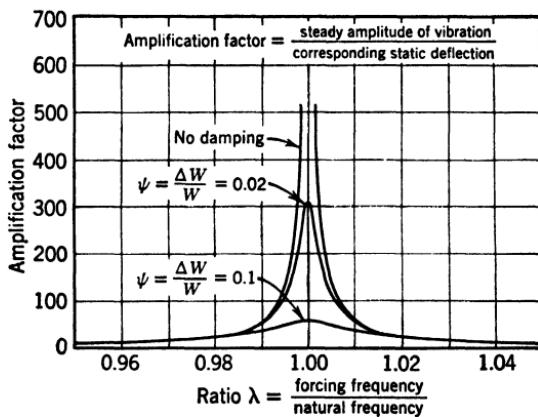


FIG. 52·2 Resonance curves for different values of specific damping  $\psi$ .

of a forced vibration as a function of the ratio  $\lambda$  between the imposed frequency and the natural frequency of the specimen in the vicinity of  $\lambda = 1$  (Fig. 52·2). The width  $\Delta\lambda$  of this curve at the ordinate of 0.5 of the maximum amplitude  $A_0$  is a direct measure of the damping capacity, to which it is related by the expression:

$$\psi = \frac{2\pi}{\sqrt{3}} \Delta\lambda \quad (52 \cdot 3)$$

If damping is defined by the *phase angle* or *phase lag*  $\gamma$  of a forced vibration, this definition can be visualized by considering that during an elastic forced vibration the acting force and the resulting deformation are in phase, both passing their extreme and their zero values simultaneously. Since the vector of the deformation velocity is normal to the force vector, no work is done. If energy is dissipated in the deformation, a component of the deformation velocity falls into the direction of the force,

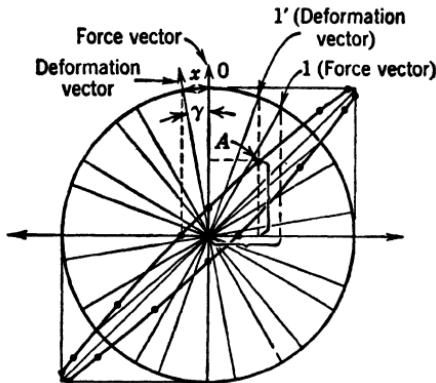


FIG. 52-3 Vector representation of damped forced vibration. Point A on damping ellipse defined by projections of pertaining force and deformation vectors.

and the deformation lags behind the acting force; the energy dissipated can be defined by the lag angle between the periodic functions of force and deformation. If deformations are represented as abscissas and forces as ordinates, one cycle of the elastic vibration is given by a straight line inclined under  $45^\circ$ , rising first toward the right, attaining a maximum value and falling towards the left and subsequently returning to the origin. In the vibration of an ideal viscous or viscoplastic material the deformation velocity is in phase with the acting force; since in a periodic vibration deformation velocity and deformation are  $\pi/2$  out of phase, there is a lag of  $\pi/2$  between the force and the deformation; the combination of vectors results in a full circle, which represents the dissipated energy. If the deformation lags only by a small angle  $\gamma$ , the pertaining momentary values of force and of deformation are located on an ellipse; this ellipse replaces the real hysteresis loop (Fig. 52-3). The introduction of the

*phase or lag angle*  $\gamma$  is thus equivalent to approximating the hysteresis loop of real materials, which cannot be expressed analytically, by an ellipse for which simple mathematical relations exist.

If a periodic force  $P = P_0 \sin \omega t$  is applied and produces a maximum deformation velocity  $v_0$ , the rate of energy loss,

$$\dot{W}_D = \frac{1}{2}P_0v_0 \sin \gamma \quad (52 \cdot 4)$$

Since the relation between velocity  $v_0$  and deformation  $A_0$

$$v_0 = \omega A_0 = 2\pi n A_0 \quad (52 \cdot 5)$$

eq. 52·4 becomes

$$\dot{W}_D = \frac{1}{2}\omega P_0 A_0 \sin \gamma \doteq \frac{1}{2}\omega P_0 A_0 \gamma = \pi n P_0 A_0 \gamma \quad (52 \cdot 6)$$

for small values of  $\gamma$ .

The energy loss per cycle,

$$W_D = \pi P_0 A_0 \gamma \quad (52 \cdot 7)$$

the maximum potential energy stored per cycle,

$$W = \frac{1}{2}P_0 A_0 \quad (52 \cdot 8)$$

Therefore,

$$\psi = 2\pi\gamma = 2\delta = \frac{2\pi}{\sqrt{3}} \Delta\lambda \quad (52 \cdot 9)$$

which is the relation between the various characteristics used to define the damping of a material.

Frequently the mechanical vibrations of the specimen are analyzed in terms of an analogy with the steady-state response of a linear electric circuit; the graphical representation in Fig. 52·3 is interpreted accordingly by considering the two axes of coordinates as the axes of the real and the imaginary part, respectively, of the complex response to the acting electromotive force  $E = E_0 e^{i\omega t}$ . The response is defined by the *impedance*  $Z(i\omega)$  of the circuit, which is the complex ratio between the electromotive force  $E$  and the current  $I$  in the steady state; it is expressed in the form,<sup>52·1</sup>

$$I = \frac{E}{Z(i\omega)} \quad (52 \cdot 10)$$

and represents a generalization of the concept of the electrical resistance to an alternating current. Because of the analogy of the differential equations governing mechanical and electric oscillations of one degree of freedom, respectively,

$$m \frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + ax = P = P_0 e^{i\omega t} \quad (52 \cdot 11)$$

and

$$L \frac{d^2Q}{dt^2} + R \frac{dQ}{dt} + \frac{1}{C} Q = E = E_0 e^{i\omega t} \quad (52 \cdot 12)$$

according to which the inductance  $L$  is equivalent to the mass  $m$ , the resistance  $R$  to the damping  $\beta$ , the reciprocal value of capacitance  $C$  to the elastic (spring) constant  $\alpha$ , the impressed electromotive force  $E$  to the mechanical force  $P$ , and the current  $I = dQ/dt$  to the rate of displacement or deformation, the concept of *mechanical impedance* can be introduced to define the steady-state response of the oscillating mechanical system to an imposed period force  $P_0 e^{i\omega t}$ . This mechanical impedance  $Z(i\omega)$  is interpreted either as a force versus displacement or force versus displacement-velocity response function. In the first interpretation it represents a generalization of the spring constant, in the second it is a generalized viscosity coefficient.

If a sinusoidal force  $P(t) = P_0 \sin \omega t$  is imposed, the resulting deformation is given by

$$x(t) = \frac{P(t)}{Z(i\omega)} = P_0[A \sin \omega t + B \cos \omega t] \quad (52 \cdot 13)$$

The mechanical impedance function  $Z(i\omega)$  thus represents a complex spring constant which expresses the elastic and inelastic responses of the vibrating system. The real component  $A$  of the spring constant defines the amplitude of the deformation which is in phase with the acting force, whereas the imaginary component  $B$  gives the amplitude of the deformation which is out of phase by  $\pi/2$ ; it represents the inelastic effect on the impedance and is therefore identical with the sine of the phase angle  $\gamma$ . Hence, for small phase angles ( $\sin \gamma \sim \gamma$ )

$$2\pi B = 2\pi\gamma = \psi \quad (52 \cdot 14)$$

The application of the electrical-mechanical analogy and of

the operational methods developed in the analysis of the response of electric circuits to electromotive forces<sup>52-1</sup> to the analysis of mechanical systems subjected to imposed forces, particularly to the analysis of their inelastic response to alternating forces, has the advantage that solutions in electric circuit theory have been worked out in great detail. Thus with the aid of the concept of impedance it is relatively easy to determine the combined response of a system consisting of many elements, for instance the system represented by a material with a discrete or continuous distribution of relaxation times.

Damping is the expression of a dissipation process of applied strain energy either into heat (real damping) or into surface energy (progressive damage by disruption of cohesive bonds). The mechanism of dissipation into heat varies even for the same material with the amplitude, and it is different for materials of different structure. The fact that the applied mechanical energy is transformed into heat energy is in itself insufficient to interpret the significance of damping observations. It is in terms of the *differences* in the operating dissipation mechanisms that differences in the significance of damping can be evaluated and damping observations interpreted and correlated with mechanical behavior. Such interpretation is frequently difficult because external dissipation mechanisms exist in addition to the internal ones; they cannot always be entirely eliminated and tend to distort the real picture.

The dissipation mechanisms operating within specimens or structures under cyclic loads are:

1. External dissipation by friction within bearings, grips, and the like, and resistance of the surrounding medium.
2. Internal dissipation by place change of particles within an essentially amorphous material.
3. Internal dissipation accompanying work hardening of single crystals or polycrystalline aggregates.
4. Internal dissipation by repeated fragmentation of a spontaneously re-forming polycrystalline structure, that is, by thermal softening or recrystallization.
5. Dissipation into surface energy, usually associated with dissipation of applied energy by thermal oscillation and by slip in the course of the formation and propagation of cracks.

The observed damping characteristics are, in general, the

expression of a combination of two or more of the afore-mentioned dissipation mechanisms; the differentiation between the individual influences is usually difficult, if not impossible. However, since the damping capacity of a specimen can be rapidly determined, there have been numerous attempts to correlate damping with other mechanical properties that cannot be easily observed and measured, such as notch sensitivity, impact strength, or fatigue performance. The result of those attempts has mostly been disappointing.

The purpose of damping measurements is usually either:

1. To study changes in damping as an indication of changes in other mechanical properties of the material which are produced by the same testing conditions, or to study damping as an indication of the presence or absence of certain characteristics.
2. To determine values of the damping capacity for direct use in design of vibrating systems near the resonance range.
3. To study the internal structure and mechanical behavior of materials by studying their inelastic response to cyclic loads under which this structure remains unaffected by the test.

To achieve any one of the foregoing objectives different

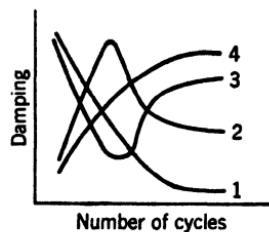


FIG. 52·4 Variation of damping with time (number of cycles).

approaches are necessary. If, for instance, damping is observed as an indication of progressive damage to cohesive strength by repeated stress cycles, it is necessary to subject the specimen in the damping and in the fatigue test to identical stress amplitudes. In this case damping is essentially the expression of changes under forced vibrations within the structure of the material; if under these conditions the damping characteristic tends towards a stable value (Fig. 52·4), this value is not associated with the material before the test, but only with the condition of the material as changed by the test. The (changing) damping values at any particular moment characterize the structure of the material only at the same moment. If different testing procedures do not produce the same permanent change in the structure of the material, or even if the sequence of changes producing the same final change is different, the test results are

not comparable, and the relations 52.9 and 52.15 are invalid. The damping capacity, therefore, cannot be an indication of any property of the material that is determined by a type of test totally unrelated to the damping test.

If the damping capacity of the material is to be determined for design purposes, this requires the setting up of testing conditions equivalent to service conditions and the determination of a final stable value of the specific damping.

Where damping observations are used to study the structure of the material, this structure must not be affected by the test. Hence, such damping tests are limited to cycles of very small strain amplitude; it is only for such conditions that the equivalence of different testing procedures and different damping characteristics may be assumed to exist.

The damping of whole structures excited into vibrations by special apparatus has occasionally been observed<sup>52.2</sup> in the expectation that the damping capacity might provide an indication of the state of the structure. However, the sources of damping in a structure are so manifold that the interpretation of a single damping constant in terms of performance seems hardly possible. Even a more or less consistent interpretation of relative changes of damping with time of service is difficult since, for instance, observations on bridges have shown tendencies of the damping both to increase and to decrease with time, although neither behavior could be reliably correlated with performance.

If external dissipation by friction in the bearings or through the resistance of the surrounding medium is eliminated, the observed damping is essentially the expression of the inelastic behavior of the material. The amount of energy that can possibly be dissipated by fracture in the creation of new surfaces is probably of a much smaller order of magnitude than the energy dissipated through the accompanying inelastic deformation. The real significance of damping under various conditions could therefore only be determined by an analysis of the effects of different testing conditions on both the general inelastic behavior and the damping values; it is reasonable to assume that, whenever an interrelation between damping and some other mechanical property can be established, this is possible only because the inelastic behavior of the material in the damping test and in the test performed to observe the particular property is nearly

identical. Hence, the possibility of an interrelation between damping and conventional static strength, or hardness or notch sensitivity or performance under impact or creep beyond the purely viscous range can be ruled out because of the widely different conditions of inelasticity and the resulting lack of correlation between the respective tests and the damping test. The only relation that may be assumed to exist is that between the performance of the material under repeated load cycles and damping as a function of the number of cycles since in this case similar testing conditions produce similarity of inelastic behavior.

Changing the testing conditions affects the inelastic behavior differently in different material; the significance of damping will, therefore, necessarily vary for various materials and conditions. The results of damping observations can therefore be interpreted only if the influence of the different factors making up the general condition of the test are individually understood and their effect on the test or performance analyzed. These factors are the same that influence inelastic behavior in general namely, stress, strain, strain rate, temperature, and the internal structure of the material as initially existing and subsequently changed in the course of its strain history.

The effect of strain rate or stress rate is usually expressed as the effect of frequency. The material in the damping test is sensitive to frequency in a similar manner as it is to strain rate or stress rate in the single-stroke test. However, there is a difference in the degree of sensitivity, since the inelasticity of the response to an external load is considerably more pronounced within certain frequency ranges of a vibration test than it is at the same stress level in the single-stroke test. Whereas at relatively low stresses the stress-strain curve of a material under a slowly applied load may show no appreciable deviation from a straight-line relation of ideal elasticity, the damping test usually reveals a definite inelastic response expressed by relatively high damping, if the duration of the imposed stress cycle is within the order of magnitude of the relaxation spectrum of the material or of its unordered phase (see Art. 37). Since the inelastic response of materials within the range of low stresses is due to energy dissipation within the unordered phase of the material, the specific damping capacity within this range of stress is a constant; the behavior can thus be roughly described by the

oscillation equation of a system of one degree of freedom with linear viscous damping. This damping is, however, an inverse function of the frequency.

Damping of one-phase or polyphase molecular materials such as polymers or glass will usually remain of the viscous type up to relatively high stresses, since the dissipation mechanism does not change with stress, unless considerable destruction of the internal network of molecular chains takes place. In polycrystalline metals, on the other hand, the dissipation mechanism depends on stress and will change considerably when the stress exceeds the level up to which no appreciable structural change takes place in the material. Below this level the deformational response is purely viscoelastic.<sup>52-3</sup> When the applied load produces changes in the structure by slip and fragmentation, the specific damping reflects both the viscous dissipation mechanism and the dissipation by crystal breakup and thermal softening. Only part of the energy expended in work hardening is dissipated into heat while part is stored up in the work-hardened structure as latent energy. Since both the extent of crystal fragmentation and the latent energy are functions of stress or of strain, or rather of both, that is, of the energy applied, the specific damping capacity can no longer be independent of stress. Observations have shown that the damping capacity in most metals increases with stress or strain amplitude in such a way that beyond the range of small amplitudes the increase becomes gradually linear (Fig. 52-5).

Grammel<sup>52-4</sup> has shown that a linear dependence between amplitude and damping (logarithmic decrement) is produced in a freely oscillating system with one degree of freedom by introducing a resistance proportional to the square of the velocity. There have been no attempts to interpret the significance of this conclusion.

At stresses at which cracks are initiated and propagated, damping expresses the dissipation of applied energy into surface energy

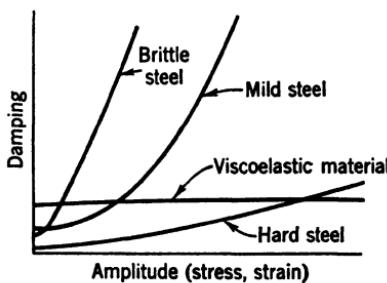


FIG. 52-5 Specific damping as a function of amplitude.

combined with the increased dissipation resulting from the intensified plastic deformation around the spreading cracks. Since the energy dissipated by plastic slip *accompanying separation* is probably a multiple of the energy actually *transformed* into surface energy, the only possibility of directly observing dissipation by surface energy would be to create conditions under which no other dissipation mechanism of similar tendency and comparable intensity would be present. Such conditions might exist in fatigue tests leading to fracture under relatively low stress levels. Observation of the change of damping with the

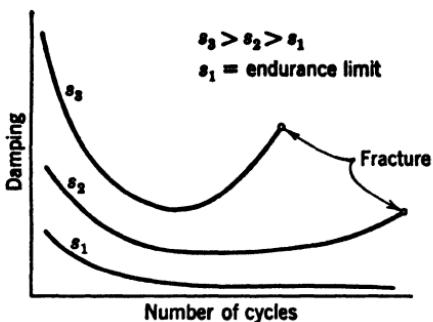


FIG. 52·6 Variation in metals of specific damping with number of load cycles for different stress levels  $s$ .

number of load cycles sustained have shown curves which, for different stress levels, are schematically reproduced in Fig. 52·6. The first part of the curves of decreasing damping is obviously the expression of the progress of the work-hardening process, the intensity of which gradually decreases until a final pattern of the crystal structure characteristic for the ap-

plied stress level is reached. The last part of the curves showing sharply increasing damping is the expression of the formation and of the spreading under considerable local deformation of macroscopic cracks, leading more or less rapidly to fracture. The central part which exists only at moderate stress is the expression of either one or the other of the two possible dissipation mechanisms or their joint effect: relaxation within the unordered and distorted regions and separation of bonds on the atomic scale accompanied by momentarily intensified thermal oscillation of the separated particles. The two superimposed effects cannot be isolated; however, the fact that the diagram in this range is not parallel to the  $N$  axis (as it would be for purely viscous damping) but rises steadily suggests a combination of both mechanism.

It may thus be inferred from Fig. 52·6 that the change of damping with the number of load cycles sustained is related to the process of progressive damage that finally produces fatigue

fracture, although it does not directly express the effect of damage, but rather the effect of localized inelastic deformation accompanying damage. It is, therefore, not the absolute value of the damping capacity, but its change with the number of cycles, which is an indication of the process of progressive damage.

Similarly, in the relation between specific damping and stress amplitude, which has occasionally been assumed to provide an indication of the endurance limit (Fig. 52-7), the amplitude, at which a definite change of the trend in this function from a direction nearly parallel to the stress axis (viscous response) to a more or less linearly increasing relation is observed, may only be considered to delimit the stress at which energy dissipation by crystal fragmentation sets in. This stress, however, would be identical with the endurance limit only if fatigue were dependent on work hardening alone, which it is not.

The damping capacity is as strongly influenced by a velocity-modified temperature as the general inelastic behavior (see Art. 19). Although the general trend of increasing damping with rising temperature holds for all engineering materials, the actual form of the relation depends on their internal structure.

The dependence of damping on the structure of the material makes the observation and measurement of the damping characteristics an outstanding tool for basic research of inelastic behavior and of structure within a range of stresses in which most other mechanical tests fail because of insufficient sensitivity. The structure of a material is studied in the vibration test in a similar manner to the way it is studied by X-ray or electron diffraction or with the aid of the microscope. Instead of observing the optical response of the structure to X rays or electron or light beams and interpreting this response, which is essentially geometrical, in mechanical terms, the observation of the inelastic response of the specimen under forced vibrations of different frequencies (vibration spectrum) evolves the mechanical response directly. Moreover, the response is not geomet-

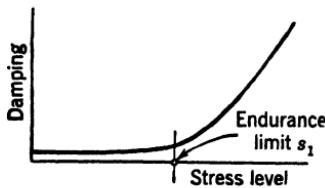


FIG. 52-7 Relation between specific damping and stress level as assumed indication of endurance limit.

rically localized, as in the optical or X-ray methods but expresses the average behavior of the particular *phase* of the material that responds to the applied frequency.

In materials in which a continuous solid phase responds elastically, the inelastic response is of the after-effect type (see Art. 37). The inelastic response is assumed to be concentrated within small local domains called *relaxation centers*, which are distributed within the elastic matrix. In metals the inelastic domains are

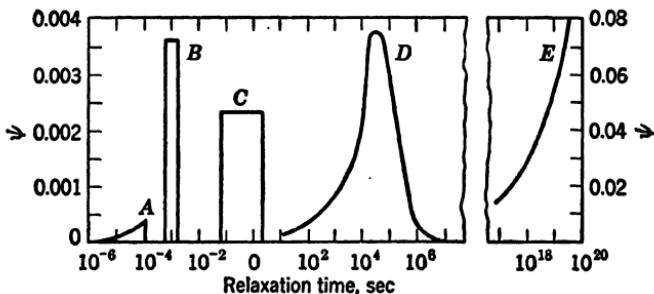


FIG. 52-8 Comparison of relaxation spectra due to different types of relaxation for iron at room temperature (after Zener<sup>52-12</sup>).

- A Relaxation by intercrystalline diffusion.
- B Relaxation by thermal diffusion in bending.
- C Relaxation by diffusion of carbon and nitrogen to and from preferred interstitial positions.
- D Relaxation within distorted slip planes and atomic layers produced by fragmentation.
- E Relaxation within grain boundaries.

identical with the grain boundaries and the distorted slip planes surrounding the crystal fragments. In amorphous materials and high polymers the relaxation (or retardation) centers are distributed within the material; they do not constitute optically identifiable components of the material, but rather different responses of individual elements (molecules) or groups defined by their individual relaxation (or retardation) times.

Experimental studies of the inelastic after effect within a range of stresses or strains which, in the usual engineering terminology, is defined as elastic, date back to Weber,<sup>52-5</sup> Kohlrausch,<sup>52-6</sup> and Voigt;<sup>52-7</sup> the theory has been developed by Boltzmann,<sup>52-8</sup> and Wiechert.<sup>52-9</sup> Within the last 30 years this effect has been extensively studied in high polymers and textiles.<sup>52-10</sup>

More recently Zener<sup>52·11</sup> and coworkers have developed an experimental approach to the analysis of the structure of metals by observation and interpretation of their *anelastic* response (relaxation spectrum), in vibration tests. The new term *anelasticity* has been introduced for the phenomenon that hitherto was known as *elastic after-effect*. However, the meaning of this term has been extended to include not only the inelastic response to stress regions of microscopic order of magnitude within the elastic matrix, such as the viscous grain boundaries and the distorted slip layers, but also the relaxation effects resulting from various diffusion processes involving relatively small numbers of individual particles such as diffusion of interstitial atoms. The positions of the individual relaxation spectra associated with the different types of relaxation involving elements of different orders of magnitude differ by several orders of magnitude, and the intensity of the effect, as expressed by both the damping and the width of the distribution of relaxation times, varies considerably (Fig. 52·8).<sup>52·12</sup>

### 53. Theory of Anelastic Effects

The theory of anelasticity is based on the principle of superposition (see Art. 38) formulated independently by Boltzmann and Wiechert and expressed by eqs. 38·1 and 38·2. According to eq. 38·1 the deformation  $x(t)$  under an applied load  $P$  is given by

$$x(t) = \alpha P [1 + \beta_c \psi_c(t)] \quad (53·1)$$

where  $\psi_c(t)$  represents the creep function and  $\beta_c$  is a constant. If this load is applied at a time  $\theta$  previous to the considered time  $t$  and removed at  $(\theta + d\theta)$ , the deformation will not disappear on load removal; the residual deformation, will only be recovered in time. Hence, the deformation at time  $t$  is equal to the instantaneous deformation due to the load  $P_t$  applied at time  $t$  plus the sum of all residual deformations due to the transient load  $P_\theta$  applied during the time interval  $d\theta$  (Fig. 53·1). According to eq. 53·1, the residual deformation at time  $t$  due to a load  $P_\theta$  acting between  $\theta$  and  $(\theta + d\theta)$  is

$$\begin{aligned} dx &= \alpha P_\theta \beta_c [\psi_c(t - \theta) - \psi_c(t - \theta + d\theta)] \\ &= -\alpha P_\theta \beta_c \frac{d\psi_c(t - \theta)}{d\theta} d\theta \end{aligned} \quad (53·2)$$

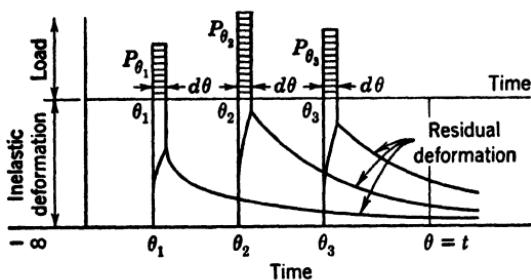


FIG. 53·1 Superposition of residual deformations produced by loads  $P_\theta$  acting at different times  $\theta$ .

The total residual deformation at time  $t$  due to the loading history  $P(\theta)$  (Fig. 53·2) is therefore the integral from  $\theta = -\infty$  to  $\theta = t$  of eq. 53·2. Hence, the total deformation,

$$x = \alpha \left[ P_t - \beta_c \int_{-\infty}^t P(\theta) \frac{d\psi_c(t-\theta)}{d\theta} d\theta \right] \quad (53\cdot3)$$

If the elapsed time is denoted by  $\lambda = (t - \theta)$ , eq. 53·3 becomes

$$x = \alpha \left[ P_t + \beta_c \int_0^\infty P(t-\lambda) \frac{d\psi_c(\lambda)}{d\lambda} d\lambda \right] \quad (53\cdot4)$$

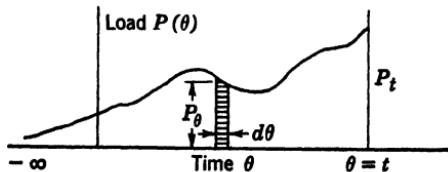


FIG. 53·2 Schematic representation of loading history diagram.

Since the creep function  $\psi_c(t)$  represents the delayed deformation, the function  $d\psi_c(t)/dt$  represents the creep rate. With  $\frac{d}{dt} \psi_c(t) = \phi_c(t)$ , eq. 53·4 may be written in the form,

$$x(t) = \alpha \left[ P(t) + \beta_c \int_0^\infty P(t-\lambda) \phi_c(\lambda) d\lambda \right] \quad (53\cdot5)$$

which is the form of the superposition principle developed by Becker.<sup>53·1</sup>

According to eq. 53·5, the residual deformation  $dx$  at a time  $t$ , due to the transient load  $P_\theta$  applied between  $\theta$  and  $(\theta + d\theta)$ , is given by

$$dx = \alpha P_\theta \beta_c \phi_c(\lambda) d\theta \quad (53 \cdot 6)$$

or, for simple loading and unloading with a constant load  $P$ ,

$$\frac{dx}{dt} = \pm x_0 \phi_r(t) \beta_c \quad (53 \cdot 7)$$

where  $x_0$  denotes the instantaneous deformation under the applied load;  $\phi_r(t)$  is a function of time, which has been called by Boltzmann the *remembrance function*. The plus sign applies if  $P_0 = 0$  for  $t < 0$  and  $P = P_0$  for  $t > 0$ , whereas the minus sign applies if  $P = P_0$  for  $t < 0$  and  $P = 0$  for  $t > 0$ .

Equation 53·5 expresses the deformation  $x(t)$ , produced by a given load  $P(t)$ . By a similar derivation, using eq. 38·2, the expression connecting the load  $P(t)$  necessary to produce a specified deformation  $x(t)$  is obtained in the form,

$$P(t) = \frac{1}{\alpha} \left[ x(t) - \beta_r \int_0^\infty x(t - \lambda) \phi_r(\lambda) d\lambda \right] \quad (53 \cdot 8)$$

The *remembrance function*  $\phi_r(t) = \frac{d}{dt} \psi_r(t)$  is the relaxation rate.

This is the form of the superposition principle as developed by Boltzmann. When the superposition principle is valid, the knowledge of the remembrance function is sufficient to compute all inelastic effects.

In the most general case of anelastic behavior of nonmetals and metals, the response of the retardation or relaxation centers is usually represented by a continuous distribution  $F(\tau)$  of retardation or relaxation times (see Art. 38), where the ordinate  $F(\tau)$  represents the contribution to the observed inelastic response—for instance the deformation in the creep test, the load in the relaxation test, or the impedance in the vibration test—of those elements the retardation or relaxation times of which lie within the range  $\tau$  and  $(\tau + d\tau)$ . Since the distribution function frequently covers a range of several orders of magnitude, it is expedient to adopt the logarithmic scale for  $\tau$  and to define the distribution in terms of the contribution of elements the retarda-

tion or relaxation times of which lie within a range ( $d \log \tau$ ) at  $\tau$ , by the distribution function  $F(\tau) d \log \tau$ . With this modified distribution the remembrance function  $\phi(t) = \frac{d}{dt} \psi$  is obtained with reference to eq. 38·10 in the form,

$$\phi = \int_0^\infty F(\tau) \tau^{-1} e^{-t/\tau} d \log \tau \quad (53.9)$$

With the aid of eq. 53·9 the relaxation or retardation spectrum can, in principle, be derived analytically if the function  $\phi_r$  or  $\phi_c$  which have been determined by experiment can be expressed as analytical functions. The rigorous procedure, however, requires the application of Laplace and Fourier transforms, since it involves the inversion of an integral. If a periodic force of frequency  $\omega$  is applied  $P(t) = P_0 e^{i\omega t}$ , the resulting deformation,

$$x(t) = \frac{1}{Z(i\omega)} P(t) \quad (53.10)$$

$Z(i\omega)$  denotes the complex steady-state mechanical impedance of the system and is defined by

$$\frac{1}{Z(i\omega)} = \alpha [1 + L(i\omega)] \quad (53.11)$$

$L(i\omega)$  is the Fourier transform of the remembrance function  $\phi(t)$ , given by the expression,

$$L(i\omega) = A(\omega) - iB(\omega) \quad (53.12)$$

where

$$A(\omega) = \int_0^\infty \phi(t) \cos \omega t dt \quad (53.13)$$

and

$$B(\omega) = \int_0^\infty \phi(t) \sin \omega t dt \quad (53.14)$$

According to eq. 52·14,  $B$  expresses the phase angle  $\gamma$  or  $1/2\pi$  of the specific damping capacity of the system; for small angles the energy dissipation per cycle is therefore  $B\pi P_0^2 \alpha$ . The amount of energy that is stored twice in every cycle is  $\alpha(1 + A)0.5P_0^2$ . In the undamped vibration the energy stored is  $0.5P_0^2 \alpha_\infty$ , where  $\alpha_\infty$  denotes the spring constant for  $\omega = \infty$ . Hence,  $\alpha = \alpha_\infty$ .

$(1 + A)$  represents the dynamic spring constant. According to eq. 53·13 for zero frequency ( $\omega = 0$ ), that is, for the creep test,

$$A(0) = \int_0^\infty \phi_c(t) dt = \psi_c(\infty) \quad \text{and} \quad B(0) = 0 \quad (53 \cdot 15)$$

where  $\psi_c(\infty)$  denotes the total creep for  $t = \infty$ . Hence, the spring constant  $\alpha$  varies between  $\alpha_0 = \alpha_\infty[1 + \psi_c(\infty)]$  for  $\omega = 0$  and  $\alpha_\infty$  for  $\omega = \infty$ , since, for infinite frequency ( $\omega = \infty$ ),

$$A(\infty) = 0 \quad \text{and} \quad B(\infty) = 0 \quad (53 \cdot 16)$$

By introducing  $\phi_c(t)$  according to eq. 53·9 into eq. 53·15, the relation is obtained,

$$A(0) = \int_0^{+\infty} F(\tau) d \log \tau = \psi_c(\infty) \quad (53 \cdot 17)$$

Thus  $A(0)$  expresses the area beneath the relaxation spectrum drawn to a logarithmic time scale.

The expressions  $A(\omega)$  and  $B(\omega)$  are convenient measures of the anelastic effects.  $A(\omega)$  defines the variation of the dynamic spring constant with the frequency of vibration, whereas  $B(\omega)$  is a measure of the variation of the specific damping.

For a material with a single relaxation time  $\tau$ , the remembrance function, according to eqs. 53·9 and 53·17,

$$\phi = A(0)\tau^{-1}e^{-t/\tau} \quad (53 \cdot 18)$$

If this expression is introduced into eqs. 53·13 and 53·14, the inelastic characteristics

$$A(\omega) = \int_0^\infty e^{-t/\tau} \cos \omega t d(t/\tau) \quad (53 \cdot 19)$$

and

$$B(\omega) = \int_0^\infty e^{-t/\tau} \sin \omega t d(t/\tau) \quad (53 \cdot 20)$$

$A(\omega)$  and  $B(\omega)$  are thus the Laplace transforms of  $(\cos \omega t)$  and  $(\sin \omega t)$ , respectively. By using tables of Laplace transforms, the relations are obtained:

$$A(\omega) = \frac{1}{1 + \tau^2 \omega^2} A(0) \quad (53 \cdot 21)$$

and

$$B(\omega) = \frac{\tau \omega}{1 + \tau^2 \omega^2} A(0) \quad (53 \cdot 22)$$

These two anelastic characteristics are plotted in Fig. 53·3 as functions of the product  $(\omega\tau)$ . For values  $(\omega\tau) > 10$ , the dynamic spring constant approaches its minimum elastic value  $\alpha = \alpha_\infty$ , whereas the damping is comparatively small, since  $A(\omega) \sim 0$  and  $B(\omega) \sim 0.1A(0)$ ; for values  $(\omega\tau) < 0.1$  the spring constant  $\alpha$  with  $A(\omega) \sim A(0)$  approaches its maximum value  $\alpha = \alpha_\infty[1 + A(0)]$ , whereas the damping with  $B(\omega) \sim 0.1A(0)$  is again small. Only in the intermediate frequency range  $0.1 < \omega\tau < 10$  are the

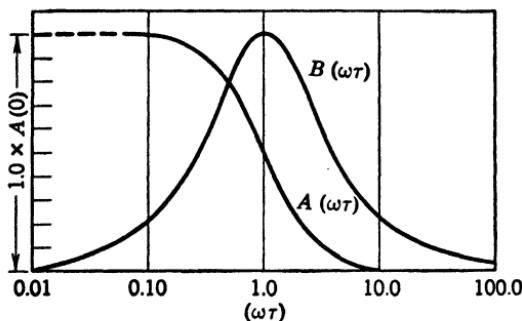


FIG. 53·3 Characteristics  $A(\omega\tau)$  and  $B(\omega\tau)$  as functions of  $(\log \omega\tau)$  for material with a single relaxation time.

variations with frequency of the spring constant and the damping appreciable. The peak of the damping function  $B(\omega\tau)$  is obtained by differentiation of eq. 53·22.

$$\frac{d}{d(\omega\tau)} B(\omega\tau) = 0 \quad (53 \cdot 23)$$

which results in the relation,

$$\tau\omega = 1 \quad \text{and} \quad \tau = 1/\omega \quad (53 \cdot 24)$$

With the aid of eq. 53·24 the mean relaxation time of materials can be evaluated from vibration tests; for polycrystalline alpha iron the relaxation time associated with grain-boundary relaxation has been estimated by Ké<sup>53·2</sup> to be of the order of magnitude of  $\tau = 0.75$  sec at a temperature of  $470^\circ\text{C}$ . Since the order of magnitude of the shear modulus  $G = 10^{12}$  dynes per sq in., the coefficient of viscosity at this temperature  $\eta = \tau G$  is of the order of magnitude of  $10^{11}$  to  $10^{12}$  poises. With the activation energy  $Q = 85,000$  cal per mole determined by experiment, the tempera-

ture dependence of the relaxation time and of the coefficient of viscosity of the grain boundaries is through the factor  $e^{42,500/T}$ . Hence, at room temperature  $\eta$  is of the order of magnitude of  $10^{48}$  poises and the mean relaxation time practically infinite ( $10^{30}$  years).

Under the assumption that  $F(\tau)$  is a slowly varying function of its arguments, the validity of eqs. 53·21 and 53·22 can, with rough approximation, be extended to the case of a continuous distribution of relaxation times. Unless this assumption can be considered justified, the anelastic characteristics  $A(\omega)$  and  $B(\omega)$  must be determined by introducing eq. 53·9 into eqs. 53·13 and 53·14. These equations thus contain an unknown distribution function  $F(\tau)$  and therefore cannot be solved rigorously. It has been suggested<sup>53·3</sup> however that the validity of eqs. 53·21 and 53·22 be extended to a continuous distribution of relaxation times by direct integration. Hence,

$$A(\omega) = \int_0^{\infty} \frac{F(\tau)}{1 + \omega^2\tau^2} d\tau \quad (53 \cdot 25)$$

$$B(\omega) = \int_0^{\infty} \frac{F(\tau)\omega\tau}{1 + \omega^2\tau^2} d\tau \quad (53 \cdot 26)$$

The experimental determination of the variability of the anelastic characteristics  $A(\omega)$  and  $B(\omega)$  from resonance vibration tests is difficult; variation of the resonant frequency through a wide range would involve considerable difficulties because the dimensions of the specimen would have to be varied very widely. Since, however, anelastic effects and temperature are interrelated through the heat of activation  $Q$ , as expressed by the relaxation-time-temperature relation 48·4, it may be expected that the variation of the anelastic characteristics  $A(\omega)$  and  $B(\omega)$  with the frequency of vibration at constant temperature can be related to their variation with temperature at constant frequency. Observation of the anelastic characteristics as functions of temperature may thus replace their observation as functions of frequency. The relation between frequency and temperature variations of the anelastic characteristics have been derived and extensively used in experimental work by Zener and coworkers.<sup>53·4</sup>

According to eq. 53·22 the damping  $B$  is a function of the

parameter ( $\omega\tau$ ). Considering the relaxation-time-temperature equivalence (eq. 48·4) for a material of unchanging structure, the expression for damping may be written in the general form:

$$B(\omega\tau) = \text{const } f(\omega e^{Q/RT}) \quad (53 \cdot 27)$$

Hence, instead of observing  $B(\omega\tau)$  at constant temperature,  $B(T)$  or  $B(1/T)$  can be observed at a constant frequency. Similarly, curves  $A(T)$  or  $A(1/T)$  at constant frequency replace the curves  $A(\tau\omega)$ . In order to derive the functions  $A(\omega\tau)$  and

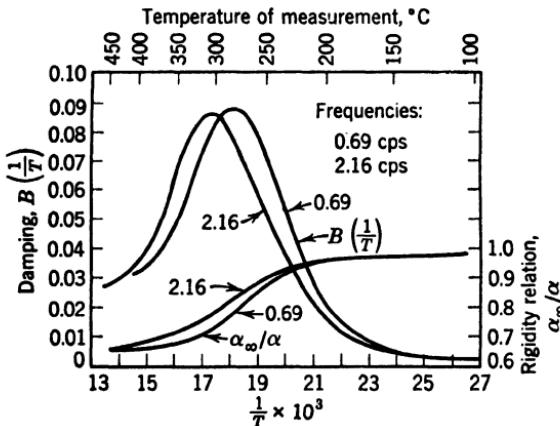


FIG. 53·4 Shift of damping characteristics as a result of changed frequency of test (after T. S. Ké<sup>63·5</sup>).

$B(\omega\tau)$  from the functions  $A(1/T)$  and  $B(1/T)$  the heat of activation  $Q$  must be determined. This can be done by two series of observations of curves  $A(1/T)$  and  $B(1/T)$  for two different frequencies. It is evident from eq. 53·27 that an increase of  $\omega$  shifts the curves  $A(1/T)$  and  $B(1/T)$ , toward higher temperatures, without changing the shape of the curves, as illustrated by Fig. 53·4; similarly, an increase of temperature shifts the curves  $A(\omega\tau)$  and  $B(\omega\tau)$  toward lower values of ( $\omega\tau$ ). By relating the relative shifts in  $1/T$  required to bring to coincidence the pair of curves  $A(1/T)$  or  $B(1/T)$  observed at two different frequencies, with the ratio of those frequencies, an equation is obtained for the determination of the activation energy  $Q$ .

The relation between frequency and temperature is derived from the condition that the shape of the damping function be

independent of temperature, or

$$\frac{dB}{d(1/T)} = 0 \quad (53 \cdot 28)$$

By differentiating eq. 53·27,

$$C \frac{d}{d(1/T)} f(\omega e^{Q/RT}) \cdot e^{Q/RT} \cdot \left( \frac{d\omega}{d(1/T)} + \omega \frac{Q}{R} \right) = 0 \quad (53 \cdot 29)$$

the relation is obtained,

$$\frac{d\omega}{d(1/T)} = -\omega \frac{Q}{R} \quad \text{or} \quad \frac{d\omega}{\omega} = -\frac{Q}{R} d(1/T) \quad (53 \cdot 30)$$

Equation 59·30, integrated, gives the relation between frequency and temperature:

$$\frac{d \log \omega}{d(1/T)} = -Q/R \quad (53 \cdot 31)$$

By replacing the differentials by the differences between  $\omega_2$  and  $\omega_1$  and between  $T_2$  and  $T_1$ , eq. 53·31 may be written in the form:

$$Q/R = \frac{\log (\omega_2/\omega_1)}{1/T_1 - 1/T_2} = \frac{\log (\omega_2/\omega_1)}{\Delta(1/T)} \quad (53 \cdot 32)$$

The activation energy  $Q$  can thus be determined from two curves  $B(1/T)$  observed at two different frequencies  $\omega_1$  and  $\omega_2$  and shifted horizontally by the amount  $\Delta(1/T)$ , so as to coincide.

In polycrystalline metals a change in grain size has a similar effect on the curves  $A(1/T)$  and  $B(1/T)$  that a change in the frequency of vibration has. When the frequency is kept constant, an increase of grain size shifts both curves towards smaller values of  $(1/T)$ , or higher temperatures  $T$ , as indicated in Fig. 53·5. Whereas the maximum amount of damping, indicated by the peak of the damping curve  $B(1/T)$ , decreases appreciably with increasing grain size, the value  $A(0)$  which defines the differences between the spring constants ( $\alpha_0 - \alpha_\infty$ ) remains independent of grain size.<sup>53·5</sup>

The effect of grain size on the anelastic characteristics is an expression of the fact that the thermal instability of the distorted atomic arrangement within the grain boundaries and slip

bands as well as the volume of distorted material increase with decreasing grain size. Thus, the smaller the grain size, the lower the temperature at which the same amount of damping due to the viscous response of the grain boundaries develops, particularly if the grain refinement is the result of work hardening. This conclusion is confirmed by Fig. 52·8 which indicates that the relaxation times within the distorted regions produced by work hardening are considerably shorter than the relaxation times of the initial grain-boundary regions. On the other hand,

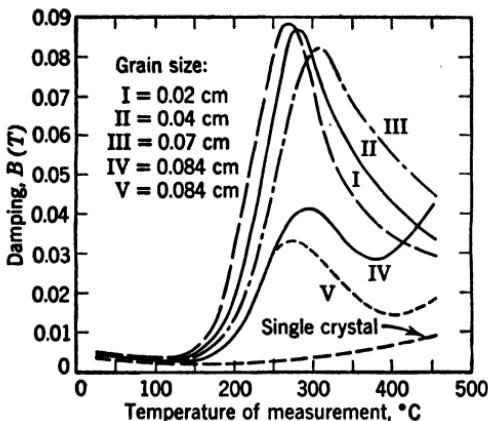


FIG. 53·5 Shift of damping curves  $B(T)$  with grain size (after T. S. K $\ddot{\text{e}}$ <sup>53·5</sup>).

the larger the grains, the less the amount of damping produced within the reduced volume of unordered and distorted domains of the material. Therefore the characteristic anelastic effects are absent in single crystals (Fig. 53·5).

The variation of anelastic damping with frequency or with temperature can be interpreted in terms of *thermal currents* within the material instead of relaxation times of structural units. Such interpretation has first been suggested by Zener,<sup>53·6</sup> who assumes that fluctuations in stress give rise to fluctuations in temperature in the same way that local fluctuations in temperature give rise to local fluctuations in stress. Temperature gradients cause irreversible thermal currents associated with damping. If these currents are able to maintain temperature equilibrium within the field of textural stresses, the vibration proceeds isothermally with little dissipation of energy. In the

other extreme case the vibration is adiabatic, and the internal dissipation of energy is again very small. Hence damping reaches a maximum only when the vibration is partly isothermal and partly adiabatic. Isothermal conditions are produced if the duration of the load exceeds that of the long relaxation times, whereas adiabatic conditions are associated with load duration much shorter than the short relaxation times, since the thermal currents are produced by the relaxing structural units which, dissipating part of their energy, emit this energy into heat.

#### 54. Impact

The effects of the applied strain rate on the inelastic behavior of materials is expressed by eq. 19·15, from which relations between strain rate and stress at constant temperature and between strain rate and temperature at constant stress can be derived. There are, however, two additional effects which are not contained in those relations but which influence the mechanical behavior of materials, particularly metals, strained beyond the elastic range at very high rates of strain. These are (1) the change from essentially isothermal conditions at low and moderate strain rates to adiabatic conditions at very high strain rates, and (2) the velocity of propagation of the plastic deformation.

When the material undergoes plastic deformation adiabatically, the stress associated with a certain deformation is raised by the high strain rate and by work hardening, as under isothermal conditions; however, because of the rise in temperature produced by the heat which is developed during the inelastic deformation and not carried away, the velocity of deformation at a certain stress level is higher under adiabatic than under isothermal conditions. The greater the strain, the greater the amount of heat developed and the higher the associated rise in temperature; the more pronounced, therefore, the increase in the velocity of deformation. The stress-strain curves for adiabatic conditions are therefore usually flatter than those obtained under isothermal conditions although the stress at which plastic deformation is initiated is considerably higher under impact than under static loading<sup>54·1</sup> (Fig. 54·1). The flattening of the curves will be the more marked, the larger the share of distortional energy within the total strain energy, since the rise in temperature depends on the amount of plastic deformation and thus on the

distortional energy alone, whereas the volumetric part of the energy, if it produces expansion, may even reduce the rise in temperature. The shear stress versus shear strain curve will therefore be flatter than the stress-strain curve in uniaxial tension. At a certain strain rate it may even become horizontal or drop after attaining a maximum, with the result that the metal flows at constant or even at decreasing stress or, if the stress is kept constant, at increasing strain rate (instable flow). Zener and Holloman,<sup>54-2</sup> by punching steel plates at extremely

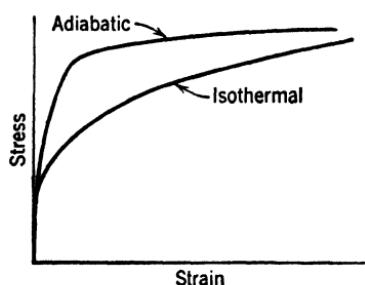


FIG. 54·1 Schematic representation of isothermal and adiabatic stress-strain curves of metals.

high strain rates and by metallurgical analysis of the metal in the vicinity of the punch, have shown that very high temperatures are attained around the punch; hence, by concentration of distortional strain the deformational resistance of the metal is substantially reduced within the small region to which, as a result of this reduction, the deformation is confined. Whereas for punching at low and moderate strain rates, the plastic deformation extends over a comparatively large region around the punch, this deformation is essentially confined to the immediate vicinity of the punch in the high-velocity punching. Therefore, the resistance to penetration, expressed in applied energy, is substantially reduced. The phenomenon of the localized reduction of resistance to deformation by the adiabatic conditions produced by very high strain rates is of primary importance in the study of the penetration of armorplates as well in the study of speed effects in cutting and machining processes.

A solution of the problem of the propagation of plastic deformation is important for the interpretation of the behavior of bars and members subject to longitudinal impact, particularly of the performance of metal specimens in tension impact tests.

Elastic strains are propagated within the strained material with the velocity of sound  $c_0$ ; considering the one-dimensional problem, the stress wave travels along the infinitely long bar without change of shape. The stresses produced in a bar, one

end of which is put into longitudinal motion with a velocity  $v_0$  are  $s = eE$ . Since  $v_0 = ec_0$ , the stress,

$$s = \frac{v_0}{c_0} E \quad (54 \cdot 1)$$

By introducing the known relation,

$$c_0 = \sqrt{E/\rho} \quad (54 \cdot 2)$$

where  $\rho$  denotes the mass density of the material, the equation results,

$$s = \rho v_0 c_0 \quad (54 \cdot 3)$$

which gives the stress produced in an infinitely long and thin elastic bar or rather wire by longitudinal impact of velocity  $v_0$ . This stress, however, is only produced within the distance  $x < c_0 t$ ; for  $x > c_0 t$  the stress,  $s = 0$ .

In a material with a nonlinear stress-strain relation the velocity of the strain wave is no longer constant, and the progressing stress wave suffers a continual change of shape. The basic theory of the propagation of plastic strain has been developed independently by Th.v. Kármán,<sup>54·3</sup> and G. I. Taylor.<sup>54·4</sup>

For a material with a curved stress-strain diagram the stress front in the specimen can be thought of as a large number of small stress increments superimposed one on the other. Each increment travels along the specimen with a speed determined by its position in the stress-strain relation. Since the slope of the relation decreases with increasing stress, every stress increment travels more slowly than the one just below it, and faster than the one above it, as can be inferred from eq. 54·2 by replacing  $E$  by  $ds/de$ . This results in a steady lengthening of the stress wave as it progresses. Although the wave front proceeds in accordance with the shape of the loading part of the stress-strain diagram, the back of the wave is governed by the elastic (unloading) part of the diagram. Therefore the back of the wave, traveling with the highest velocity of the wave front, will gradually overtake the parts of the wave front, which move with smaller velocities. Along the distance over which this takes place the stress will be constant, decreasing beyond it; within this distance the irrecoverable strain is also constant.

The equation of motion of an element of a one-dimensional elastic bar, hit longitudinally by an impact load, has the form:

$$\frac{\partial^2 u}{\partial t^2} = c_0^2 \frac{\partial^2 u}{\partial x^2} = c_0^2 \frac{\partial e}{\partial x} \quad (54 \cdot 4)$$

where  $u$  denotes the longitudinal displacement in the direction of  $x$ . v. Kármán has extended the validity of eq. 54·4 to bars of a material with a general stress-strain relation  $s = s(e)$  by introducing the tangent modulus  $T = ds/de$  into eq. 54·3 instead of the constant elastic modulus. Hence,

$$\rho \frac{\partial^2 u}{\partial t^2} = T \frac{\partial^2 u}{\partial x^2} = \frac{ds}{de} \cdot \frac{\partial e}{\partial x} \quad (54 \cdot 5)$$

The velocity  $c_p$  of propagation of plastic strain at a certain value  $e_n$ ,

$$c_p = \sqrt{\frac{T_{(e=e_n)}}{\rho}} \quad (54 \cdot 6)$$

is therefore a variable, depending on the slope of the stress-strain diagram at the point  $e = e_n$ . The solution of the differential equation has to fulfill the boundary conditions  $u = v_0 t$  for  $x = 0$  and  $u = 0$  for  $x = \infty$ . If it is assumed that the strain  $e$  is a function of  $x/t$  alone, the complete solution of eq. 54·5 is

$$\begin{aligned} & \text{for } x < v_1 t && e = e_1 \\ & \text{for } v_1 t < x < ct && T(e) = \rho x^2/t^2 \\ & \text{for } x > ct && e = 0 \end{aligned} \quad (54 \cdot 7)$$

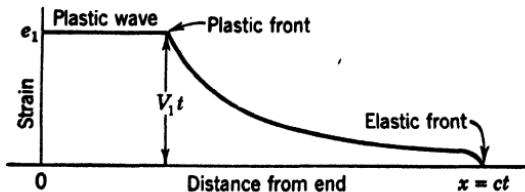


FIG. 54·2 Distribution of strain of plastic wave as a function of distance  $x$  from impacted end (after v. Kármán<sup>54·3</sup>).

The distribution of  $e$  as a function of  $(x/t)$  is shown in Fig. 54·2. The velocity of the plastic wave  $v_1$  and the maximum

strain  $e_1$  as a function of the velocity of impact  $v_0$  is obtained from the condition that the displacement of the point  $x = 0$  is  $v_0 t$ ; hence, considering eq. 54·6,

$$v_0 = \int_0^{e_1} (T/\rho) de \quad (54\cdot8)$$

where  $T = f(e)$ . Equation 54·8 determines  $e_1$  as a function of the imposed impact velocity  $v_0$ .

The stress in the bar varies from 0 beyond the elastic wave front, over a gradually increasing range up to  $s_1$  within the plastic region; the stress  $s_1$  corresponds to  $e_1$ . The strain  $e_1$  which is a function of  $v_1$  is obtained from eq. 54·8 by graphical computation, integrating the curve  $\sqrt{(ds/de)/\rho}$  versus  $e$ , derived from the experimental stress-strain curve. If it is assumed that  $ds/de \rightarrow 0$  for large values of  $e$ , the integral 54·8 has a maximum which might be considered a *critical velocity*; the strain  $e_{1,\max}$  associated with this velocity would then be the strain at which an unstable condition leading to rapid fracture is reached in a test.

The principal assumptions on which the present theories of the propagation of plastic waves are based are that the stress-strain relation is independent of time and that the process is isothermal, so that the tangent moduli of the static stress-strain curve may be used in the evaluation of eq. 54·8. These assumptions may be justified as a first approximation for very thin metal wires only. For bars of finite cross section transversal heat flow and radiation will considerably affect the deformation process, making the use of the isothermal static stress-strain curve inadequate. The critical velocity actually observed in impact tests of bars and associated with a maximum value of energy dissipated during fracture (Fig. 54·3) thus cannot be satisfactorily interpreted in terms of isothermal deformational

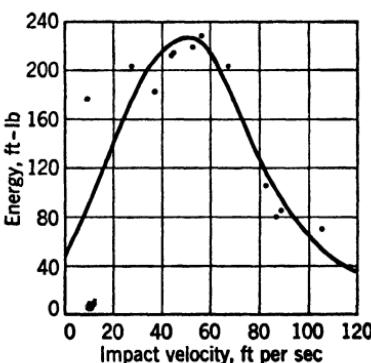


FIG. 54·3 Variation for hard copper of energy to fracture with impact velocity (after Hopmann<sup>54·6</sup>).

behavior alone, using the nominal or "engineering" stress-strain diagram instead of that of true stress and logarithmic strain under adiabatic conditions.

It has already been pointed out that the difference between the stress-strain diagram observed under static (isothermal) and extreme dynamic (nearly adiabatic) conditions is of primary importance in the analysis of all deformation and fracture phenomena at high strain rates, such as that of metal cutting and milling. As a result of the time and temperature effects the

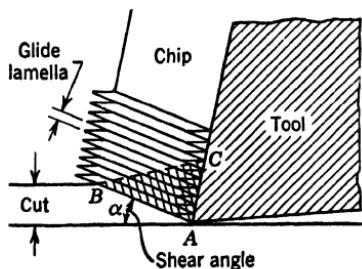


FIG. 54·4 Shear zone before tool tip in metal cutting (schematic).

applied at bullet velocities. Assuming a moderate cutting speed of 100 ft per min, a depth of cut of 0.01 in., a shear angle  $\alpha$  of about  $20^\circ$  ( $\sin \alpha = 0.34$ ), a thickness of the glide lammellas of the order of magnitude of  $10^{-4}$  cm (see Art. 18) and a shear strain of the order of magnitude 0.1 associated with the deformation at the yield stress, the strain rate of the shearing process is of the order of magnitude of  $(100 \times 12 \times 2.54 \times 0.34 \times 0.1)/(60 \times 10^{-4}) = 15,000$  per sec. At such a strain rate, which is by several orders of magnitude larger than the strain rates applied in conventional tension tests, the yield stress must be a multiple of the static yield stress (see eq. 19·13). However, the heat developed in the course of the plastic deformation in the region of shear before the tool tip will tend to reduce the isothermal dynamic yield stress, this reduction being the more pronounced, the more nearly adiabatic the shearing process. It has been found<sup>54·6</sup> that the actual yield stress at strain rates of the order of magnitude of the above estimated rate of shearing is however still more than three times the static yield stress.

deformational response of the metal in the shear zone before the tool tip (Fig. 54·4) is substantially different from the response under static conditions and varies with the cutting speed. It can be shown by a rough calculation that the strain rates imposed in the cutting process are extremely high and exceed even the strain rates produced in tension or compression tests by impact loads

In the cutting process the heat developed before the tool tip is the less, and this heat is the more effectively carried away, the smaller the depth of the cut and the smaller the cutting speed. Therefore the specific cutting force which is directly proportional to the yield stress will increase with decreasing depth of the cut and with decreasing speed. On the other hand, the harder the material to be cut, the higher the required cutting speed. Since under stationary conditions the shearing takes place at constant stress, all the work-hardening effects having been eliminated by the increase of the yield stress and by the heat effect, the stress-strain diagram of the ideal plastic body appears to be a fairly good approximation of the real deformational behavior of metals under extremely high strain rates, if the appropriate dynamic yield stress is considered instead of the static stress. Hence, problems of plastic deformation at very high strain rates could be analyzed more effectively by applying the theory of ideal plastic flow after the dynamic yield stress has been exceeded, than by attempting to adapt methods used in the solution of elastic problems, introducing a variable tangent modulus instead of the constant modulus of elasticity. The cutting problem, for instance, can be analyzed by assuming a triangular ideal plastic zone ( $ABC$  in Fig. 54·4) in which the shearing process is concentrated and in which, therefore, the transition takes place from the yield stress in the shearing surface before the tool tip, to zero stress at the beginning of the chip.<sup>54·7</sup>

### References

- 52·1 L. A. PIPES, *Applied Mathematics for Physicists and Engineers*, McGraw-Hill Book Co., New York (1946) 141.
- 52·2 R. K. BERNHARD, *Mechanical Vibrations*, Pitman Pub. Corp., New York (1943).
- 52·3 R. WEGEL and H. WALTHER, *J. Applied Phys.* **6** (1935) 141.
- 52·4 R. GRAMMEL, *Z. Ver. deut. Ing.* **58** (1914) 1600.
- 52·5 W. WEBER, *Pogg. Ann. Phys.* **34** (1834) 247; **24** (1841) 1.
- 52·6 R. KOHLRAUSCH, *Pogg. Ann. Phys.* (3) **12** (1847) 393.
- 52·7 W. VOIGT, *Ann. Physik* **47** (1892) 671.
- 52·8 L. BOLTZMANN, *Ann. Physik* **7** (1876) 624.
- 52·9 E. WIECHERT, *Ann. Physik* **50** (1893) 335.
- 52·10 H. LEDERMAN, *Elastic and Creep Properties of Filamentous Materials and Other High Polymers*, Textile Foundation, Washington, D. C. (1943).

52·11 C. ZENER, *Elasticity and Anelasticity*, Chicago Univ. Press (1948) 60.  
52·12 C. ZENER, *Metals Technol.* (1946) T.P. 1992.  
53·1 R. BECKER, *Z. Physik* **33** (1925) 185.  
53·2 T. S. KÉ, *Metals Technol.* (1948) T.P. 2370.  
53·3 T. ALFREY, *Mechanical Behavior of High Polymers*, Interscience Pub., New York (1948) 193.  
53·4 T. S. KÉ, *J. Applied Phys.* **20** (1949) 274.  
53·5 T. S. KÉ, *Phys. Rev.* **71** (1947) 553.  
53·6 C. ZENER, *Phys. Rev.* **53** (1938) 90.  
54·1 S. D. CLARK and G. DATWYLER, *Proc. ASTM* **38** (1938) 98.  
54·2 C. ZENER and J. H. HOLLOWOM, *J. Applied Phys.* **15** (1944) 22.  
54·3 TH. v. KÁRMÁN, *Natl. Defense Research Council A-29; Office Sci. Research & Development* **365** (1942).  
54·4 G. I. TAYLOR, *J. Inst. Civil Engrs.* **26** (1946) 486.  
54·5 W. H. HOPPMANN, *Proc. ASTM* **47** (1947) 541.  
54·6 G. I. TAYLOR, *loc. cit.* 502.  
54·7 E. H. LEE and B. W. SHAFFER, *Rept. 43, Grad. Div. of Applied Math. Brown Univ.*, Providence (1949).

## CHAPTER

## 11

## FRACTURE

**55. Theories of Brittle Fracture**

It has been pointed out in Art. 22 that the theoretical cohesive strength of solids, derived from the interatomic forces is between a hundred and a thousand times larger than the observed values. In order to explain this discrepancy it is assumed that the interatomic forces across a potential fracture surface are not overcome simultaneously but in a certain sequence; a crack is thus initiated under an over-all force which, by producing separation of a limited number of bonds only is but a small fraction of the force that would be required to break all bonds simultaneously. This assumption presupposes a highly nonuniform distribution of bond strength (that is, of bond energy) within the material. Such distribution is actually the result of the inhomogeneous structure of real materials as well as of the existence of high textural stresses. The discrepancy between the observed fracture strength of a real material and the computed cohesive strength of the ideal homogeneous atomic lattice is therefore a necessary consequence of this inhomogeneity and imperfection of the real structure (see Art. 13).

As long as the material is assumed to be homogeneous, the explanation of the low cohesive strength requires the introduction of the concept of *stress concentrations* of different severity as a consequence of which the stresses within the material are magnified locally by a factor adequate to raise the peak stress at certain points to the order of magnitude of the theoretical separation strength. As the most obvious cause of such stress

magnifications the existence within the unstressed material of cracks of varying size and shape has been postulated, and a concept of crack propagation as well as a statistical analysis of crack distribution and density has been developed on the basis of which the cohesive strength of a specimen of finite dimensions can be derived from the theoretical separation strength or from the surface energy of the separation surfaces. However, it is obvious that the structure of real materials is not homogeneous; since it is therefore not necessary to assume the existence of cracks in order to account for the discrepancy between the observed and the theoretical fracture stress, the numerous *micro-crack theories*<sup>55·1</sup> should not be considered as attempts to reproduce real conditions, but rather as attempts to develop a theory of fracture consistent with the concept of the isotropic homogeneous ideal material, the behavior of which is governed by the theory of elasticity.

Statistical theories of fracture have mainly been derived under the assumption of brittle fracture; hence, their applicability is very much restricted. Under conditions of homogeneous stress for which the statistical microcrack theories have been developed, engineering materials generally do not fracture by clean separation, neither preceded nor accompanied by inelastic deformation. Therefore the fact that, whenever fracture is associated with inelastic deformation, the structure of the material at the end of the fracture process is not identical with the structure at its start invalidates the usual statistical approach to fracture; this approach is based on an assumed *unchanging* initial distribution of microflaws, microcracks, or defects which cause stress concentrations of such severity that, under a load producing an over-all stress equal to the actually measured fracture stress, the local stress at the severest microcrack reaches the intensity of the theoretical cohesive strength of a material of perfectly homogeneous atomic structure.

Essentially, statistical theories of fracture in one form or another are all based on Griffith's theory of crack propagation<sup>55·2</sup> in which the strain energy necessary to break an elastic solid containing an elliptical crack is computed. The solution of the elastic problem of the narrow transversal elliptical hole (Fig. 55·1) within the homogeneously stressed field, which was first

derived by Inglis,<sup>55-3</sup> gives the maximum stress  $s_{\max}$  at the edge of the hole,

$$s_{\max} = 2s_0 \sqrt{c/\rho} \quad (55 \cdot 1)$$

where  $s_0$  denotes the mean stress,  $\rho$  the smallest radius of curvature, and  $2c$  the length of the ellipse. On the basis of Inglis' solution Griffith computed the difference  $\Delta W_e$  between the elastic energies that can be stored up in a very thin homogeneously stressed specimen without and with an elliptic hole. This difference is given by the expression,

$$\Delta W_e = \frac{1}{8}\pi c^2 s_0^2 \quad (55 \cdot 2)$$

The surface energy that appears with the formation of a crack of length  $2c$  is given by

$$W_s = 4c\sigma \quad (55 \cdot 3)$$

where  $\sigma$  denotes the specific surface energy. The total change in potential energy resulting from the formation of a crack of length  $2c$  is therefore

$$\Delta W = \Delta W_e - 4c\sigma \quad (55 \cdot 4)$$

The condition for an instability which would result in the extension of the crack of length  $c$  is

$$\frac{\partial \Delta W}{\partial c} = 0 \quad (55 \cdot 5)$$

By differentiating eq. 55·4 and considering eq. 55·2, the relation is obtained,

$$s_0 = \sqrt{\frac{2E\sigma}{\pi c}} \quad (55 \cdot 6)$$

$s_0$  represents the over-all stress under which a crack of length  $2c$  spreads. Griffith has tested eq. 55·6 by breaking, under internal air pressure, glass tubes and hemispheres with cracks of length between 0.15 and 0.89 in. produced by cutting with a

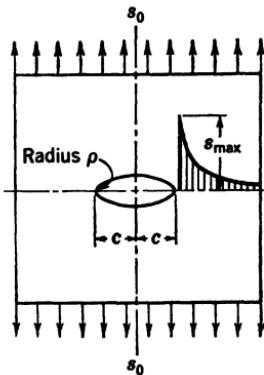


FIG. 55·1 Stress distribution adjacent to elliptical hole (after Inglis<sup>55-3</sup>).

diamond. He found the constancy of the product  $s_0 \sqrt{c}$ , predicted by eq. 55·6, well confirmed.

If a breaking stress  $s_0$  is attributed to every size of crack within a homogeneous stress field, or a *strength*  $s_0$  to every homogeneously strained element containing one crack of this size, and it is assumed that a material body is made up of volume elements each of which contains one crack, the assumption of a distribution, within the body, of cracks of various sizes is equivalent to the assumption of a distribution, within the body, of the strength of the volume elements containing such cracks. Different values of stress  $s_0$  are therefore required to produce fracture in the individual elements containing cracks of different sizes. If the cracks are distributed at random with a certain specific density, and if it is assumed that there is no interrelation between the individual processes of crack propagation, the strength of the entire specimen will actually be determined by the weakest volume element, or the lowest "strength," that is, by the lowest value of local strength associated with the volume element containing the most severe crack. Thus, Griffith's theory implies that, if the stress  $s_0$  is exceeded within a certain volume, the resulting instability at the end of the crack with the most critical stress concentration is sufficient to ensure the propagation of the crack throughout the entire volume. Hence, the statistical microcrack theories of fracture are based on the concept of the *weakest link*, according to which the strength of a chain (being a model of the considered specimen) is determined by the smallest value of the strength of a link to be found in a sample of size  $n$ , where  $n$  is the number of chain links in the model (or the number of cracks or defects in the specimen). In other words it is assumed that the most severe crack or defect determines the fracture strength of the entire specimen..

If, at constant crack density, the volume of the specimen increases, the total number of cracks equally increases and so does the probability of encountering a severe crack. A relation can therefore be established between the strength of a specimen defined by its largest crack or weakest volume element and the volume of the specimen, if the distribution of cracks of different severity in the material is known and can be expressed by the distribution of (volume) strength values.

In terms of statistical theory the problem is thus one of dis-

tribution of extreme values, that is, of finding the distribution function of the smallest value of strength as a function of the number  $n$  of cracks for a given distribution function of crack severity or sizes or of strength values  $s_0$ . Different forms of the distribution functions of strength or of crack size have been assumed by different investigators, the most frequent being the Gaussian or normal distribution<sup>55·4</sup> and the exponential distribution.<sup>55·5</sup>

The assumption of a Gaussian distribution is not so self-evident as it has been assumed to be by the investigators proposing it and as might be inferred from the consideration that this distribution is assumedly that of pure chance. In the case of cracks or defects, however, the probability of a very large defect is not the same as that of a very small defect; this fact precludes the application of distribution functions which are symmetrical with regard to their mean values. It is more consistent with observations to assume that the probability of occurrence of a defect decreases in proportion with the size of the defect, from a small probability for a very large defect to an appreciable probability for a very small defect. The probability distribution of size of defects  $p(c)$  may therefore be more adequately expressed by the simple exponential or Laplace distribution,

$$p(c) = ae^{-ac} \quad (55\cdot 7)$$

where  $c$  is a measure of the size of the crack and  $a$  a parameter of the distribution, than by the symmetrical Gaussian distribution. Introducing this distribution of crack size  $c$  into eq. 55·6, the pertaining distribution of strength  $s_0$  is obtained.

The analysis of the weakest-link problem is well known in applied statistics and is of importance in various fields of science and technology. Its solution is based on the so-called *asymptotic theory* developed by Fisher and Tippett,<sup>55·6</sup> which is concerned with the derivation of the distribution function of extreme values for large samples from the given distribution function of the statistical "population." A summary of this theory has been presented by Cramer.<sup>55·7</sup>

If the distribution of strength values  $s$  of the population is given by the function  $p(s)$ , the associated cumulative distribution function is

$$P(s) = \int_0^s p(s) ds \quad (55\cdot 8)$$

The distribution of the smallest value in samples of size  $n$ , drawn from the population, is given by the probability distribution,

$$\pi_n(s) = np(s)[1 - P(s)]^{(n-1)} \quad (55 \cdot 9)$$

or by the associated cumulative distribution function,

$$\Pi_n(s) = \int_0^s \pi_n(s) ds = 1 - [1 - P(s)]^n \quad (55 \cdot 10)$$

The most probable value of the smallest value in samples of size  $n$  is represented by the *mode* of the distribution function  $\pi_n(s)$  which can be found either graphically, by drawing the function  $\pi_n(s)$  for different values of  $n$ , or analytically, by solving the equation  $\frac{d}{ds} \pi_n(s) = 0$ . If  $p(s)$  is a Laplace distribution, it can be shown<sup>55·8</sup> that the modes of the distribution of the smallest values, which represent the most probable strength values of specimens containing  $n$  cracks, decrease linearly with  $(\log n)$ ; since  $n = \gamma V$  where  $\gamma$  denotes the crack density, that is, the average number of cracks or defects per unit volume, and  $V$  the volume of the specimen, the dependence of strength  $s$  on volume  $V$  is of the form,

$$(s_1 - s_2) = \text{const} (\log V_2 - \log V_1) \quad (55 \cdot 11)$$

For a Gaussian distribution of crack or flaw strength  $s_0$ , the strength of specimens has been found<sup>55·4</sup> to decrease linearly with increasing  $\sqrt{\log V}$ . The assumption concerning the distribution of crack size contained in eq. 55·7, on the other hand, leads to the conclusion that the strength of specimens of volume  $V$  decreases as the reciprocal of  $\sqrt{\log V}$ .<sup>55·8</sup>

Solutions such as eq. 55·11 which give only the dependence of the *most probable value* of the smallest values are not sufficient, if it is desired to establish the actual *frequency distribution* of the smallest values. This distribution can be derived in terms of a new variable,

$$r = nP(s) \quad (55 \cdot 12)$$

which, introduced into eq. 55·9, gives the expression for the distribution function  $g_n(r)$  for sample size  $n$ :

$$g_n(r) = \left(1 - \frac{r}{n}\right)^{n-1} \quad (55 \cdot 13)$$

The asymptotic development of this function for the number of cracks or  $n \rightarrow \infty$  is given by

$$g(r) = \lim_{n \rightarrow \infty} g_n(r) = e^{-r} \quad (55 \cdot 14)$$

Equations 55·12 and 55·14 can be used to determine the frequency distribution of the strength of specimens, containing large numbers of cracks or flaws. For a Laplace distribution of flaw

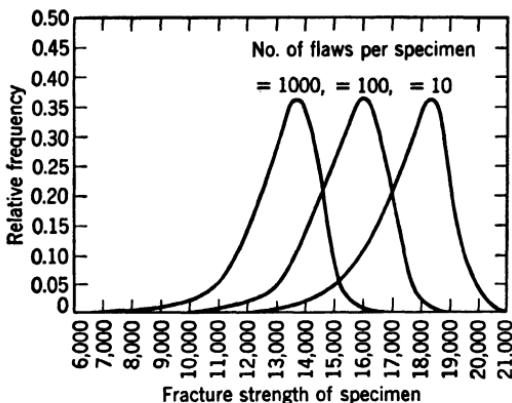


FIG. 55·2 Frequency distribution of strength of specimens containing various numbers of flaws distributed according to the Laplace distribution  
 $p(s) = 5 \times 10^{-4} e^{-(0.001s - 20)}$  (after Epstein)<sup>55·8</sup>.

strength, for instance, the frequency distributions  $\pi_n(s)$  have been established for different numbers of flaws (Fig. 55·2). These frequency distributions are characteristically skew, with negative skewness values. The skewness, which is actually observed in tests of the strength of materials under embrittling conditions, is the necessary result of the application of the statistical theory of the weakest link, that is, of the distribution of smallest values in large samples, to the phenomenon of brittle strength.

Because of the underlying assumptions of perfect brittleness of the material, the practical applicability of statistical theories of strength is limited to the analysis of size effects with respect to the strength of materials which essentially are brittle or which are strained under conditions in which they appear brittle. An additional limitation of the applicability of these theories is that

the conditions under which the weakest-link concept, derived from the Griffith theory, can be assumed to hold are not frequently met.

It has been pointed out that statistical theories of fracture are derived on the assumptions that (a) the flaws or cracks are independent and thus do not affect each other, (b) one single crack or defect of critical size is sufficient to produce a rapidly spreading fracture of the specimen extending over its entire cross section, no matter what its size and (c) the large-scale elastic concept of stress concentration around an elliptical hole in an isotropic medium may be extended to the analysis of the propagation of cracks of microscopic or submicroscopic size within a discontinuous structure. The first two assumptions can be justified for amorphous or single-crystal specimens of small cross-sectional area when each length element of the specimen may be considered a *chain link*, since only then would an incipient crack propagate rapidly through the whole section. It is in strength tests of thin brittle threads or in tests of statistically isotropic brittle specimens with cracks of macroscopic size that the weakest-link theory may therefore be expected to lead to a correct prediction of size effects. For specimens of large cross section with submicroscopic cracks these assumptions, however, are not justified.

The validity of the assumption that the elastic concept of stress concentration around an elliptical hole may be extended to describe the response of the atomic or molecular structure is rather doubtful. The confirmation of Griffith's theory obtained on specimens with a single crack of macroscopic size in itself does not justify the extension of this concept to the submicroscopic structure by reducing the size of the crack to atomic dimensions. The concept of the elastic stress concentration has a meaning only in terms of the continuous elastic medium; on the level where the material is considered to be built up of discrete particles it necessarily loses its reality.

It is, however, possible to base the statistical approach to brittle fracture directly on the assumption of the existence of a statistical distribution function of the separation strength of atomic or molecular bonds within the material built up of discrete particles, without recurrence to the concept of defects or of stress concentrations. This assumption, which appears justified

in view of the inhomogeneity of the potential field of the interatomic or molecular bond forces, is basically different from the assumption of microcracks and stress concentrations, since it implies that the distribution of bond forces or of bond energies is not, like that of stress concentrations, static and invariable but rather of the nature of a continually fluctuating distribution, as the waves of thermal energy pass through the material. Thus, instead of the probability of bond separation at a particular location, or of an invariable spatial distribution of bond strength, the probability is introduced of the coincidence of the separation strength of an individual atomic or molecular bond with such a force in this bond produced by the external load, which is sufficient to overcome the momentary level of bond strength at the unspecified location at which this coincidence occurs, thus producing a crack of atomic dimensions.

If the frequency distribution of the separation strength  $s$  of atomic bonds is given by the function  $p_1(s)$ , the cumulative distribution function is given by  $P_1(s) = \int_0^s p_1(s) ds$ . Under an external load  $S$ , causing a macroscopically homogeneous state of stress, but a microscopically and submicroscopically inhomogeneous field of forces, the forces  $\sigma$  induced in the statistically oriented individual bonds must also be considered as a distribution function of bond forces  $p_2(\sigma)$ . The probability that an individual bond is destroyed as the result of the action of the external load is thus the product of the probability of occurrence in the bonds of forces  $\sigma$ , by the respective probabilities that the bond strength  $s < \sigma$ , as required by the condition of coincidence defined previously. Such a combination of the distributions  $p_2(\sigma)$  for different levels of the external load  $S$  expressed in terms of the mean value of the bond force  $\sigma_0$ , with the cumulative distribution function  $P_1(s)$  results in a new cumulative distribution curve of the strength of atomic or molecular bonds  $P(S)$  expressed in terms of the external load  $S$  or the mean value of the bond force  $\sigma_0 = S/m_0$ , where  $m_0$  is the total number of bonds over the resisting section. The probability distributions  $p_1(s)$  and therefore  $p(S) = d/dS [P(S)]$  may be assumed to be of the Laplace type, since the probability distribution of separation strength of atomic bonds can be expected to be a function that decreases with increasing bond strength.

It is this concept of the bond strength varying as a result of the thermal-energy fluctuations in the material that makes it possible to extend the statistical approach to brittle fracture under repeated load,<sup>55-9</sup> since it implies that the probability of separation of an individual atomic bond by the action of the external load  $S$  is different for each load application; it also changes with the duration of the load cycle. If  $P$  is the probability that a bond will be disrupted during one load cycle of amplitude  $S$  and very short duration,  $(1 - P)$  is the probability that it will not be disrupted; the probability that it will not have been destroyed after  $N$  load cycles is  $(1 - P)^N$ . Hence, the probability that the bond will not sustain  $N$  cycles without disruption becomes  $[1 - (1 - P)^N]$ . If in a single load cycle the simultaneous severance of  $m$  of the  $m_0$  bonds would result in a total separation of the considered specimen of the material, because the simultaneous disruption of these  $m$  bonds would so considerably increase the forces in the remaining bonds that these would be rapidly destroyed, the probability that  $N$  load cycles will not cause such separation, because all  $m$  bonds remain undamaged, is  $(1 - P)^{mN}$ ; hence, the probability of separation is  $\Pi = [1 - (1 - P)^{mN}]$ . The form of  $\Pi(S)$  is identical with  $\Pi_n$

in eq. 55·10; the pertaining frequency distribution  $\pi(S) = \frac{d}{dS} \Pi(S)$

has therefore the same form that is characteristic for the functions drawn in Fig. 55·2. Hence, the differences between the modes of the individual distribution functions for different values ( $mN$ ) are proportional to the difference of  $(\log mN)$ . Thus the most probable, that is, the expected strength values of specimens of  $m$  bonds subjected to  $N$  load cycles are related by the equation:

$$\begin{aligned} S_{N_2} - S_{N_1} &= -k(\log mN_2 - \log mN_1) \\ &= -k(\log N_2 - \log N_1) \end{aligned} \quad (55 \cdot 15)$$

The *fatigue strength*  $S$  under  $N$  load repetitions is therefore expressed by the relation,

$$S = S_0 - k \log N \quad (55 \cdot 16)$$

where  $S_0$  denotes the "static" or single-stroke brittle strength.

If a Gaussian distribution of bond strength is assumed instead of the Laplace distribution, the brittle fatigue strength decreases linearly with  $\sqrt{\log N}$  instead of  $\log N$ .

Equation 55·16 is in fair agreement with experimentally established relations  $S(N)$  within the finite life range of many materials (Fig. 55·3). Where systematic deviations from this relation occur, they are due to the modifying effect of inelastic deformation, as well as to the fact that the weakest-link concept is inadequate to explain the progress of fracture in specimens of moderate or large size.

In order to develop a statistical theory of fracture, free of the inadequacy of the weakest-link concept, the model of the single chain from which this concept is derived should be replaced by a model consisting of a bundle of parallel chains. In this model the strength of the bundle does not depend on the strength of the weakest link in a single chain, but on the distribution of weakest links in all chains, and on the probability of such a coincidence of weakest links in the chains that, under a certain

load, a process of progressive fracture of chains is initiated by the consecutive breaking of individual chains overloaded as a result of preceding chain fractures. The statistical analysis of this composite chain model is considerably more complex than that of the weakest-link model, since under a certain load the probability of fracture of an individual chain is not independent of the sequence of the fractures preceding that of the considered chain. In a recent study of the strength of bundles of  $n$  parallel threads<sup>55-10</sup> under a load  $S$  producing initially forces in the threads  $S/n$ , it was found that the bundle would not break if and only if the strength of  $k$  among the threads exceeds  $S/k$ , where  $k \leq n$ .

The problem of the distribution of strength of bundles of  $n$  threads is in this case no longer a problem of the distribution of extreme values. Therefore the distribution of the strength of bundles obtained in this investigation is no longer skew but asymptotically normal for large values of  $n$ . Probably even the fracture of brittle real materials cannot be adequately described

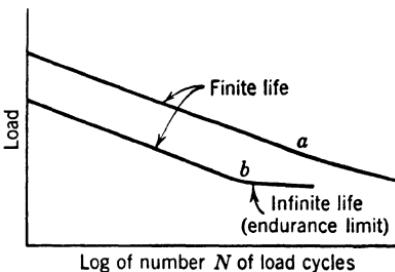


FIG. 55·3 Characteristic shape of  $S$ -log  $N$  diagrams representing results of fatigue test for materials (a) without and (b) with endurance limit.

by either the weakest link or the parallel bundle theory. These theories indicate, however, the progress of fracture under the idealized limiting conditions.

Recently an attempt has been made to derive the observed brittle strength of isotropic materials from the melting heat of the material,<sup>55-11</sup> by assuming that melting constitutes a breaking due to the heat movement of the atoms. The fracture strength would therefore be connected with the melting energy rather than with the sublimation energy. Although by this assumption the right order of magnitude of the brittle fracture strength is obtained from considerations of atomic processes without the necessity of introducing imperfections or inhomogeneities of the atomic lattice, its justification is rather doubtful.

## 56. Ductile and Brittle Fracture

In practically all engineering materials fracture is preceded and accompanied by irrecoverable deformation of one or a number of the constituent phases of the material. The deformed material at the end of the fracture process has therefore undergone considerable structural changes. In metals these changes are associated with the creation within the distorted atomic arrangement of the fragmented structure of unstable, that is, of overstrained bonds; in amorphous materials they are the expression of the redistribution and concentration of the response to the external forces in a gradually decreasing number of bonds due to bond relaxation, associated with the place change of particles connected by bonds of low stability.

Although separation and relaxation of unstable high-energy bonds are different phenomena, their damaging effect under sustained load is cumulative. Since the applied load must be balanced by an increase of the potential energy of the interatomic bonds while, at the same time, certain bonds are relaxed by place change of particles, an increased force is transferred into the remaining unrelaxed bonds. This increase intensifies the tendency to bond disruption by increasing the potential energy that can be gained by the propagation of a crack. Thus, the potential damage within the regions of unordered atomic or molecular arrangement, produced by the gradual concentration of the response to the external forces within a decreasing number of bonds of increasing energy content leads to progressive fracture,

which is preceded and accompanied by inelastic deformation; this deformation expresses the extent of place changes of particles during the load application.

Hence, within materials of unordered atomic structure the process that controls progressive damage is the same process that controls inelastic deformation. Therefore the relation of fracture strength versus time for such materials, or for polyphase materials under such conditions that the deformation of the unordered or distorted atomic regions is the fracture-controlling process, should be essentially of the same type as the strain versus time relation of the associated inelastic deformation. The dependence on stress of the rate of progressive fracture and of the rate of inelastic deformation, respectively, would thus be of the type of eq. 49·1. For sustained constant stress this equation leads to a linear relation between the logarithm of time to fracture and a function of stress which decreases with increasing stress. This relation can be derived by integrating the rate of progressive fracture, that is, the rate of growth of the fracture surface  $dA/dt$ , expressed by

$$\frac{dA}{dt} = \text{const } \frac{de_c}{dt} = \text{const } \sinh(s/s_0) \quad (56\cdot 1)$$

over the time  $t_0$  required to produce fracture by total separation.

Fracture occurs after the concentration of the response to the applied load by relaxation and bond disruption has reduced the resisting area to such an extent that it is no longer able to carry the momentary stress. This minimum area  $A_c$  is a direct function of the initial stress  $s$  since it can be assumed that the stress intensity at which the progressive fracture process is terminated by sudden separation is a constant. Hence,

$$\frac{A_{0s}}{A_c} = \text{const} \quad \text{and} \quad \frac{A_c}{A_0} = \text{const } s \quad (56\cdot 2)$$

The difference between the initial area  $A_0$  and the minimum area  $A_c$  represents the progressively disrupted area. Hence,

$$\begin{aligned} \frac{A_0 - A_c}{A_0} &= \frac{1}{A_0} \int_0^{t_0} \frac{dA}{dt} dt = \text{const} \int_0^{t_0} \sinh(s/s_0) dt \\ &= 1 - \text{const } s \end{aligned} \quad (56\cdot 3)$$

and

$$\text{const } \sinh(s/s_0)t_0 = (1 - \text{const } s) \quad (56 \cdot 4)$$

For low stresses eq. 56·4 can be approximated by

$$\text{const } (s/s_0)t_0 = (1 - \text{const } s) \quad (56 \cdot 5)$$

or

$$\begin{aligned} t_0 &= C_1(s/s_0)^{-1} - C_2 = C_1[(s/s_0)^{-1} - (\bar{s}/s_0)^{-1}] \\ &= C_1 \frac{1}{s} (1 - s/\bar{s}) \quad (56 \cdot 6) \end{aligned}$$

where  $\bar{s}$  denotes the fracture stress for  $t_0 = 0$  under the assumption that the same mechanism of deformation and of fracture is responsible for fracture at all times between  $t_0 = 0$  and  $t_0 \rightarrow \infty$ . Since this assumption is usually not justified as the mechanism of very short-time fracture is different from that of long-time fracture, the validity of eq. 56·6 is limited to values of  $s$  small in comparison with  $\bar{s}$ , so that the term  $s/\bar{s}$  can be neglected in comparison with unity. Hence, within this range eq. 56·6 can be transformed into a straight-line relation in double-logarithmic scale:

$$\log t_0 = \text{const} - \log s \quad (56 \cdot 7)$$

For large stresses eq. 56·4 converges toward

$$\text{const } e^{(s/s_0)}t_0 = (1 - \text{const } s) \quad (56 \cdot 8)$$

Relation 56·7 is valid under conditions in which the deformation is essentially concentrated within the unordered phases of the material and the time to fracture is long, whereas relation 56·8 applies if the main part of the deformation that precedes fracture is caused by slip and fragmentation of crystals, and time to fracture is short. Hence, creep-fracture test results for steel at moderately elevated temperatures and relatively low stresses when fractures appear brittle can be fairly well represented by straight lines in double-logarithmic representation, whereas test results for lead or for steel at elevated temperatures within the range of high stresses when fractures appear ductile approach straight lines in semilogarithmic representation.

Frequently the observed fracture-strength-time curves consist of ranges of different slope (Fig. 56·1). This is an indication

of a transition of the character of the creep fracture and the preceding deformation either from the intercrystalline brittle to the transcrystalline ductile type, or from this latter type to an essentially intercrystalline type in which the deformation preceding fracture is associated with extensive recrystallization and fracture itself is associated with precipitation phenomena and high-temperature corrosion phenomena, such as oxidation.

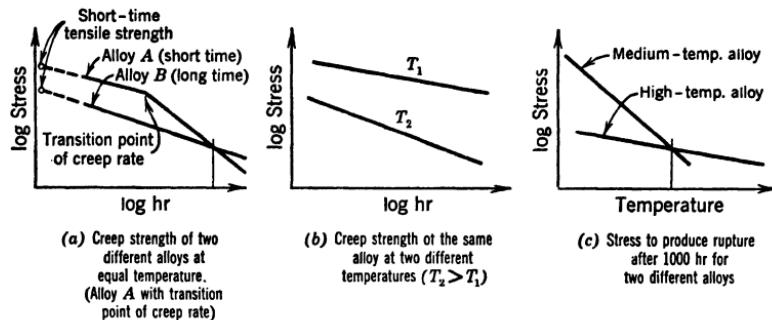


FIG. 56-1 Creep strength of heat-resistant alloys as function of time or temperature.

If on the basis of the previous considerations the relation of the rate of creep fracture versus temperature is assumed to be of the same form as the relation between creep rate and temperature, the fracture rate can be expressed by

$$\frac{dA}{dt} = \text{const } f(s)e^{-Q/RT} \quad (56 \cdot 9)$$

Hence, the fracture condition 56-3 leads to the relation,

$$f_1(s)e^{-Q/RT_0} = \text{const} \quad (56 \cdot 10)$$

or the time to fracture at a definite level of sustained stress,

$$\log t_0 = C_1 + C_2/T \quad (56 \cdot 11)$$

This relation between temperature and creep life at constant stress is confirmed by tests.

On the basis of the same assumption the damaging effect of a temperature history  $T(t)$  can be evaluated by solving the equation,

$$\int_0^t e^{-Q/RT(t)} dt = e^{-Q/RT_0} \cdot t_0 \quad (56 \cdot 12)$$

for the time to fracture  $t$  at a constant stress and temperature history  $T(t)$ , where  $t_0$  denotes the time to fracture under the same stress at a constant temperature  $T_0$ . The damaging effect of a sequence of  $n$  different temperature levels  $T_i$  sustained during respective periods  $t_i$  will thus reduce the time to fracture according to the relation,

$$t = t_0 - \sum_1^n t_i e^{-\left(\frac{1}{T_i} - \frac{1}{T_0}\right)Q/R} \quad (56 \cdot 13)$$

which represents an approximate law of "cumulative damage" for the superposition of the effects of various temperature levels in creep fracture tests.

The damaging effect of relaxation of bond forces is characteristic of the condition of materials under sustained load. In the unloaded state place changes of particles and the resulting dissipation of the excess bond energy represent a process of equalization of the energy levels of particles and thus of their forces of interaction; the result is a more uniform distribution of the response of all bonds to subsequently applied external forces. Relaxation of unstable bond forces has therefore a double aspect:

1. It is beneficial as long as it occurs spontaneously, that is, in the unloaded condition or during unloading; because of the resulting equalization of the inhomogeneities in the potential field of the bond forces, it increases the over-all level of applied stress at which fracture is initiated on the atomic scale.

2. It may be damaging or beneficial when it occurs under sustained load; by gradually concentrating the response to the applied forces within a decreasing number of bonds, it produces a "chain-reaction" process of bond separation on the atomic scale and thus increases the rate at which cracks propagate.

However, this intensification of the separation process is a trend which develops only if the period of load application is considerably longer than the lowest order of magnitude of relaxation times of the deforming material. The response to loads the duration of which is shorter or of the same order of magnitude may be quite different. In this case the relaxation of the bonds between particles of lowest stability and thus of shortest relaxation time produces at first a beneficial relief of the bond forces of highest intensity without appreciable effect of redistribution. It is at a later stage only that bonds of higher stability and longer

relaxation times start to relax, producing concentration of the response to the external force in a decreasing number of bonds. Thus, it depends on the ratio between the duration of the load cycle and the order of magnitude of the shortest relaxation times whether the effect of relaxation during load application is beneficial or damaging.

Relaxation of bond forces within the unordered or distorted material is the most important single factor that determines the character and rate of the fracture process. Only if this process is very rapid, as in fracture under rapidly applied or impact loads, appreciable relaxation effects cannot develop.

Fracture within the crystalline regions of metals may be of the brittle type, if slip is prevented either by a high critical shear stress within the eligible slip planes or by a stress system inhibiting inelastic deformation. In this case separation proceeds along the cleavage planes of the crystals; otherwise, fracture will be preceded and accompanied by extensive slip and fragmentation.

Fracture always starts locally, by submicroscopic crack formation in the direction normal to the largest separation distance of particles. However, the change in the local stress system around the cracks which form and advance into the crystal regions may be sufficient to initiate slip, wherever the state of hydrostatic tension at the advancing edge of the crack is reduced or wherever favorably oriented slip planes are encountered. The resulting local increase in the apparent ductility of the metal around the edge of the crack is bound to produce substantial slip. At points where the slip movement is blocked by crystal boundaries, a state of hydrostatic tension develops which is again overcome by the opening of a small crack perpendicular to the direction of the largest separation which, being the result of slip, approaches the direction of principal shear.

Thus the initial short clean crack normal to the direction of the largest principal strain is changed into an oblique rough crack spreading in the general direction of principal shear; it is repeatedly interrupted by steplike short clean cracks normal to this general direction. The change in character of fracture for constant temperature and strain rate depends therefore essentially on the state of stress. There is no basic difference in the fracture mechanism; the type of fracture is but the expression of the type of stress that produces it and that determines the character

and extent of inelastic deformation by which fracture is preceded and accompanied.

A change in the character of fracture as expressed by the extent of accompanying inelastic deformation may also be produced by varying either temperature or rate of load application. Since decreasing temperatures or increasing strain rates tend to raise the slip resistance along the eligible slip planes of the individual crystals, the amount of deformation within the crystal region which is associated with a certain level of applied stress decreases with increasing rate of load application or decreasing temperature. If the stress is high enough to produce extensive disruption within the intercrystalline boundaries, the rapid transfer of the response to the load into the crystal regions accompanied by the increased resistance of those regions to slip may produce such conditions within the most critically oriented crystals that cleavage-planes become operative.

### 57. Transition Temperature

Although changes in any of the three parameters: state of hydrostatic tension, strain rate, and temperature produce a gradual reduction of inelastic deformation accompanying fracture

and thus a change in the amount of energy dissipated in the course of the fracture process, these effects alone cannot account for the appreciable drop, within a relatively narrow range of temperature, in the amount of this energy that is being observed in tests of various materials (Fig. 57-1). It must therefore be assumed that a decrease of temperature within a certain range suddenly reduces either the fracture stress within the cleavage planes or the over-all stress at which fracture is initiated by the disruption

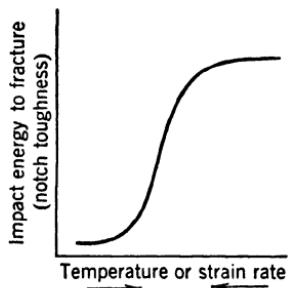


FIG. 57-1 Energy "absorption" to fracture as function of temperature or of strain rate.

of the intercrystalline regions and distorted slip planes.

There is no reason to assume an abrupt drop with temperature in the cohesion of the cleavage planes. However, a reduction of the over-all stress at which fracture is initiated within the distorted slip planes may result from such rapid increase, with

decreasing temperatures, of the relaxation times of those regions that, instead of the short-time beneficial relaxation under load of instable bonds, bond disruption would become the operative mechanism of stabilization of the position of particles. The simultaneous reduction, with decreasing temperature, of both the inelastic deformation and of the stress at which fracture is initiated is probably sufficient to account for the abruptness of the observed drop with temperature of the energy dissipated during fracture of certain metals, particularly of steel. The narrow range of temperatures in which this drop is observed is known as the *transition temperature*.

It may be assumed that every temperature  $T$  defines a distribution of relaxation times (relaxation spectrum) characterizing the inelastic response of the material at this temperature. An optimum range of duration of the load will then exist at which the relaxation of the most unstable bonds is just sufficient to increase the uniformity of the response of the bond forces to the applied load, without being extensive enough to become damaging; a shorter duration of the load would then produce less relief of inhomogeneities in the potential field of the forces between the particles and therefore a smaller resistance to the initiation of atomic separation, whereas a longer duration would produce damaging relaxation, also associated with a reduction of this resistance. The optimum range would thus define a maximum value of the resistance in terms of the over-all stress at which the fracture process is initiated within the relaxing material or the relaxing regions of a polyphase material. For loads of a duration shorter than the optimum, this stress would drop rapidly towards the limit at which no appreciable relaxation occurs; for loads of longer duration, a gradual decrease of the fracture initiation stress would result from the increasing extent of damaging relaxation.

The assumed effect of the duration of the load in relation to the relaxation spectrum is shown schematically in Fig. 57·2. For duration of the load  $t \leq \tau_0$ , where  $\tau_0$  denotes the optimum range, a comparatively small shift of the distribution curve of relaxation times towards higher values of  $\tau$  by reducing the temperature  $T$ , or by increasing the rate of load application, that is, by decreasing the duration  $t$  of the load cycle, is frequently sufficient to eliminate all relaxation effects by decreasing the

diffusion or reaction rates to practically zero. For  $t > \tau_0$  an increase of  $t$  or an increase of the temperature and the resulting shift of the relaxation spectrum toward shorter times  $\tau$  extends the relaxation to a gradually increasing number of bonds. With increasing duration of the load or increasing temperature the number of bonds with relaxation times  $\tau > t$ , which are therefore able to carry the load for a certain time, is thus reduced,

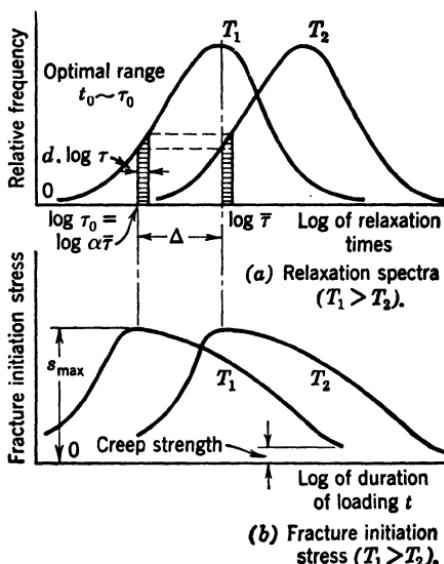


FIG. 57-2 Schematic representation of probable variation of fracture initiation stress  $s_B$  with duration of loading  $t$  resulting from distribution of relaxation times in the material.

at first very slowly and then more and more rapidly as the number of resisting bonds decreases. Hence, the over-all stress at which fracture starts within the relaxing material is reduced with increasing length of the load cycle which is the phenomenon observed in creep-rupture tests.

Inelastic deformation preceding and accompanying fracture in a material of amorphous structure increases with the duration of the load. The amount of strain energy "absorbed" or dissipated prior to fracture is obtained as the product of stress and of deformation, only the elastic component of which is proportional to the applied stress. The variation of the absorbed

energy with the duration of the load cycle is obtained by multiplying the respective values of the fracture stress by the sum of the preceding elastic and inelastic deformation, as indicated in Fig. 57·3. It is clear that the direct relation between duration of loading and inelastic deformation accentuates the abruptness of the transition from a high to a low energy absorption for loads of relatively short duration. The effect on this energy of the gradual decrease of the fracture stress with increasing duration of the load above the transition range, however, is compensated by the increasing inelastic deformation, with the result that the energy absorption above this range remains nearly constant.

With decreasing temperature the distribution function of relaxation times is shifted towards larger values of  $\tau$ ; the duration of the load at which the transition from high to low energy absorption occurs is therefore necessarily a function of temperature. In order to obtain a rough indication of the trend of this function, it is assumed that the optimum fracture stress is associated with a duration of the load  $t$  equal to a constant fraction  $\alpha\bar{\tau}$  of the mean relaxation time  $\bar{\tau}$  of the material, defined by the condition  $\log \bar{\tau} - \log \alpha\bar{\tau} = \text{const}$  (Fig. 57·2). Since the relation between relaxation time and temperature is given by eq. 48·4, the relation between temperature and load duration for the optimal fracture stress can be expressed by

$$t = \alpha\bar{\tau} = \text{const } e^{Q/RT_0} \quad (57\cdot1)$$

where  $T_0$  denotes the upper limit of the transition range, defined by the optimal value of fracture stress and therefore by the highest value of energy absorption. Hence,

$$T_0 = Q/R \frac{1}{\log t - \text{const}} \quad (57\cdot2)$$

which shows that the transition temperature of amorphous materials *decreases with increasing duration of the load*.

In metals a transition temperature from ductile to brittle fracture, that is, from fracture preceded and accompanied by a considerable dissipation of energy to fracture associated with very small dissipation, may be expected to exist under conditions for which the duration of the load is within the range of the relaxation times of the distorted regions surrounding the fragmented crystals. However, in metals the effects of precipitation

of foreign particles into and out of the distorted slip planes (see Art. 20) may considerably modify the relation between the fracture initiation stress and the duration of the applied force resulting from the effect of relaxation of bond forces by self diffusion.

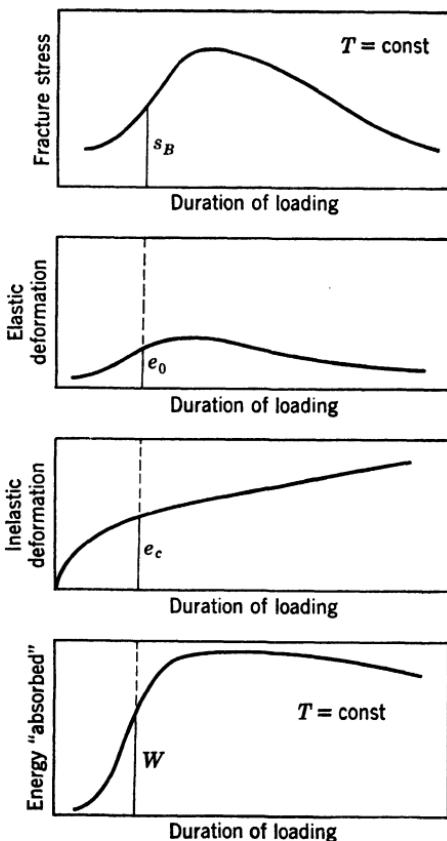


FIG. 57·3 Schematic representation of variation with duration of loading of energy "absorbed" to fracture derived from variation of fracture initiation stress and of deformation (amorphous material) [ $W \sim \frac{1}{2}s_B(e_0 + e_c)$ ].

Thus, at temperatures at which the rate of diffusion of foreign particles is of the order of magnitude of the rate of load application, the resulting unblocking of slip within the strain-aged crystalline regions may reduce the inhomogeneities within the atomic structure and consequently increase the over-all stress at which fracture is initiated. With increasing rates of load

application, that is, with reduced duration of the loading, a certain range will therefore be reached in which the fracture initiation stress drops very rapidly because slip during the period of load application remains effectively blocked. Because of the large volume of the crystalline regions in polycrystalline aggregates in comparison with that of the intercrystalline boundaries and distorted atomic layers, the blocking of slip by the effects of precipitation of foreign particles will be considerably more significant within the range of relatively short duration of the applied load than the effects of relaxation of unstable bonds by self-diffusion within the intercrystalline regions. There is, however, no difference in the trend of the variation of fracture initiation stress with duration of the load resulting from either of those effects as indicated schematically in Fig. 57·2. It is only the time range in which the drop occurs and its magnitude that may differ.

The variation, with duration of the load, of the over-all stress under which fracture is initiated either within the distorted atomic layers surrounding the crystal fragments or within the crystalline regions can thus be represented by a function of the type indicated in Fig. 57·2, in which the time scale may be related to the rate of self-diffusion (relaxation times) within the intercrystalline regions or to the rate of diffusion of foreign particles within the lattice (slip delay times). The amount of energy dissipated in the fracture process is again obtained as the product of the fracture initiation stress and the deformation. Thus, the relation between the energy dissipated in the fracture process and the duration of the applied load is of a shape very similar to the relation of the fracture initiation stress and load duration, with an accentuated drop in the abruptness of the transition from high to low energy absorption. With decreasing temperature the slip delay times are shifted toward longer values; the duration of the load at which the drop from high to low energy absorption occurs is, therefore, a function of temperature of a type roughly similar to relation 57·2, since the temperature dependence of the self-diffusion process (relaxation) is essentially of the same form as that of the diffusion of foreign particles.

As in metals the occurrence of a sharp transition from high to low energy absorption with decreasing temperature or increasing strain rate depends on the existence of slip delay effects due

to diffusion of foreign particles (strain aging), a transition temperature will probably exist only if the rate of slip is significantly affected by precipitation processes. Thus, the same alloys that are susceptible to strain aging may show a distinct transition temperature, whereas, in other metals in which a transition can be due only to the varying response of the small volume of the unordered intercrystalline regions which is of the same type as that of amorphous materials, such transition should be practically absent or very gradual.

### 58. Fracture under Single-Stroke, Sustained, or Repeated Load

Fracture occurs if the applied energy cannot be dissipated into heat nor balanced by the mobilization of additional resistance through work hardening, so that a stable condition of the loaded body can only be attained by the spreading of a crack. The fracture stress or fracture energy necessarily depends on the extent of previous deformation, because of the change of structure associated with such deformation. However, the increase of fracture stress with the extent of prior deformation might be observed only under such experimental conditions by which any further plastic deformation accompanying the fracture process of the previously deformed material could be completely blocked. Since even at very low temperatures or very high rates of applied strain, fracture of metals is preceded and accompanied by plastic deformation, these conditions cannot be fully realized. Several investigators have attempted to determine the relation between fracture stress and previous deformation,<sup>58-1</sup> using the technique of prestraining the specimens at room temperature and subsequently breaking them in liquid air, and have found a linear increase of fracture stress in tension with tensile prestrain. However, the test results are open to the objection that plastic deformation of most metals is only partly blocked by low temperatures. Actually the reduction of the limiting crystallite size with decreasing temperature (Art. 13) extends the range of crystal fragmentation under load (Art. 45), and thus tends to increase the range of deformability of the polycrystal.

Prestraining introduces anisotropy of the structure which is the more pronounced the larger the amount of strain. An attempt has been made<sup>55-1</sup> to interpret the increase of the frac-

ture stress as a result of this anisotropy, by assuming that initially randomly oriented microcracks seek a preferred orientation under the influence of inelastic strain. Simple elongation would thus tend to line the cracks in the direction of the strain, whereas in simple compression they would line up perpendicular to the axis of strain, with the result that the effectiveness of the cracks to reduce the over-all strength in tension would be reduced by pretension strains and increased by precompression strains. Although this conclusion would explain the Bauschinger effect, this effect can be equally well explained in terms of the potential energy of textural stresses produced by prestraining and stored up in the intercrystalline boundaries and distorted atomic layers (see Art. 44), without introducing the concept of microcracks.

The brittle fracture stress should be intrinsically independent of temperature.<sup>58-2</sup> On the other hand, the heat oscillation of atoms, by giving rise to local density fluctuations, may be the cause of a reduction of fracture strength with increasing temperature. However, this temperature dependence of fracture stress, which has been reported by a few investigators,<sup>58-3</sup> may also be the result of the dependence on temperature of the deformation preceding and accompanying fracture. This would explain the fact that the observed relation of fracture stress and temperature is of a similar exponential form as the relation between yield stress and temperature.

Fracture stress depends on the strain rate, since the amount of preceding plastic deformation is determined by this rate. Although the resistance to inelastic deformation is considerably less dependent on strain rate in crystalline than in amorphous materials, this dependence cannot be neglected. Hence, the higher the increase of resistance to plastic deformation due to increasing strain rate, the less additional resistance must be provided by work hardening before a definite critical fracture stress is attained, if fracture were defined by a definite value of the fracture stress. It has, moreover, been observed that even under conditions of very high strain rates the influence of preceding and accompanying deformation on the fracture process is very pronounced, particularly the effect of change from essentially isothermal to adiabatic conditions. Thus, in the fragmentation of steel shell cases the steel disintegrated at a total strain of some 35 to 40 percent. The initial velocity of the shell

fragments as well as their number were found to be a function of the strain velocity of the preceding expansion.<sup>58-1</sup>

If fracture were not essentially dependent on the preceding and accompanying inelastic deformation, it could hardly be expected that results of fracture tests were reproducible. The process of separation itself, depending essentially on inhomogeneities, both on the submicroscopic and the microscopic scale, is necessarily highly erratic. However, this process is apparently less important than the inelastic deformation preceding and accompanying it. The fact that the effect of deformation in reducing the inhomogeneities is fairly well reproducible provides an explanation of the reproducibility of the results of fracture tests, as well as the explanation of the observed increase of the scatter of test results with increasing brittleness of fracture.

So far, the considered loading conditions were such that, after a more or less rapid disruption of the apparently rigid intercrystalline and distorted crystal regions, the process of deformation and subsequent fracture would be concentrated within the crystal regions. If, on the other hand, the rate of energy dissipation within the intercrystalline and distorted crystal regions is of the same, or of a higher, order of magnitude as the rate of energy application, the deformation is concentrated within these regions, which therefore appear soft. Their cohesion is gradually destroyed by the cumulative effects of relaxation under load and separation. The deterioration of the intercrystalline cohesion under sustained or slowly applied load extends over the entire specimen since the rotation and relative motion of crystals along the intercrystalline boundaries is not localized in the regions of highest stress, as it is in the case of plastic deformation following the sudden breakdown of the grain boundaries under a rapidly applied load. The observed decrease in over-all density of the material and its appearance under the microscope suggest that the disintegration of the intercrystalline regions is followed by the appearance of microscopic cracks at or near the crystal boundaries which, during a later stage, spread into the neighboring crystal. (Fig. 49·3.)

Fracture under repeated load is the result of the combined effect of damage to the crystals and to the intercrystalline regions produced by crystal fragmentation and bond separation during the loading part, and of relaxation of potential damage (unstable

bonds, textural stresses) during the unloading part of the load cycles and during rest periods.<sup>58-6</sup> The *endurance limit* under repeated load cycles is the highest over-all stress which, at the temperature and the applied rate of strain, does not produce such disruption of the intercrystalline domains as would permit appreciable slip and crystal fragmentation to spread through the specimen. The extent of intercrystalline breakdown under a given intensity of stress determines the rate per stress cycle of the spreading of slip and fragmentation and thus the rate of formation of unstable bonds within the distorted slip bands and atomic layers surrounding the crystal fragments. The rate of bond separation under repeated stress, which defines the number of cycles a specimen can sustain prior to fracture, that is, the *life* of the specimen, depends not only on the rate of *formation* of unstable bonds, but also on the rate of their *relaxation* during the unloading part of the stress cycle. Effects which increase the total extent of relaxation, such as rest periods, without affecting the rate of formation of unstable bonds, will therefore tend to increase the life of the specimen. The effect of rest periods, however, can be appreciable only if the relaxation times of the intercrystalline region are several orders of magnitude longer than the duration of the individual load cycle. Otherwise, relaxation of the unstable bonds accompanies or follows so rapidly on the unloading part of the cycle that subsequent rest periods have no further effect; in certain metals, such as copper, relaxation effects at room temperature are so extensive that they constitute partial recrystallization and produce increased fragmentation in subsequent cycles and therefore damage, that is, a shortening of the fatigue life. Since the relaxation times are exponential functions of the activation energy, relatively small differences in the activation energy of different metals or of the same metal at different prestrains will produce different behavior in fatigue.

When the intercrystalline regions are relatively soft and deformable even under very low over-all stresses, if only the time of their application is long enough, there is no stress limit below which the spreading of slip within the crystalline areas could be restrained by the rigidity of those regions, provided the duration of the loading is sufficiently long. Damage by the formation of new unstable bonds within distorted slip bonds is therefore

produced at any level of applied stress above the critical shear stress of the largest crystal size present in the aggregate. No definite endurance limit above this stress can therefore exist. Hence, unless the intercrystalline regions are rigid and strong enough to restrain the spreading of slip, the endurance limit is a function of grain size and decreases with increasing size of the crystal grains making up the polycrystal. The endurance limit will therefore be raised by grain fragmentation associated with the work-hardening process, whereas the fatigue life may be reduced because of the creation of unstable bonds within the distorted atomic layers surrounding the fragments. The observed endurance limit will therefore be the higher above the theoretical minimum limit, associated with slip initiation in the largest crystal size, the more pronounced and the more stable the work-hardening effect in the metal. The finite life range of a metal during which it can be broken by a finite number of load cycles will therefore be the more significant compared to its endurance range, the more pronounced the effects of relaxation, recovery, and of subsequent thermal softening.

In mild steel the existence of an endurance limit considerably above the minimum limit defined by the slip resistance of the largest crystals is due to the rigidity, that is, the extremely high stability, at room temperature and below, of the intercrystalline regions. Since this rigidity also accounts for the existence of a sharp yield point, the assumption appears reasonable that, when metals manifest both a definite endurance limit and a definite yield point, the trend in the change of both characteristics with changing conditions will be parallel. This conclusion is borne out by the observation that coarse-grained steels which have a definite yield limit and relatively high endurance limit at room temperature lose both those limits at elevated temperatures. On the other hand, metals such as aluminum which have no sharp yield limit at any temperature do not have an endurance limit above that defined by the extremely low slip resistance of a single crystal. Only very fine-grained metals, in which the over-all stress producing slip is higher than the over-all stress required to disrupt the intercrystalline regions, can manifest an endurance limit without having a definite yield limit. In general, the existence of an endurance limit depends on the existence of a stable condition of the material. Where changes occur with

time, no matter whether they depend on or are independent of the load, such as creep, recrystallization, or corrosion, no endurance limit can exist.

The frequency of the repeated stress cycle can be of no effect on the fracture strength as long as the rate of energy dissipation, that is, of relaxation within the intercrystalline regions, is negligibly small compared to the rate of energy application; this is, however, no longer true when both rates are of the same order of magnitude. Under such conditions repeated load cycles will

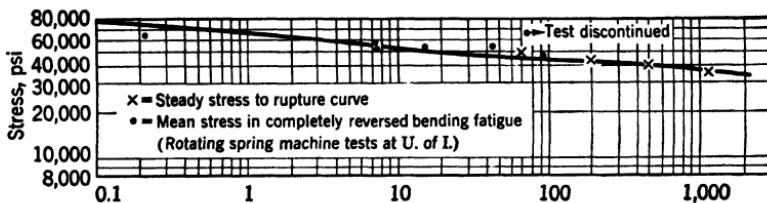


FIG. 58-1a Comparison of high-temperature creep and fracture-stress and reversed bending fatigue strength of S-590 as function of testing time (Ni-Cr-Cb-Fe alloy; compiled by R. B. Smith<sup>58-6</sup>).

have two effects: (1) The instantaneous effect of crystal fragmentation of an extent that depends on the stress level alone; this may be called the *cycle effect*. It depends on the *number* of load cycles. (2) The time-dependent effect of gradual deterioration of the cohesion of the intercrystalline and distorted crystal regions by relaxation and bond disruption, accompanied by a certain amount of work hardening by slip and fragmentation which follows the relative movement and rotation of grains along the relaxing boundaries and produces, at elevated temperature, intensified recovery and recrystallization; this is called the *creep or time effect*. It depends on stress, on temperature, and on the duration of the load cycle and, thus, for a definite number of load cycles, on the *total duration* of the test.

Both effects are always present; it is the character of the material, particularly its slip resistance and recrystallization temperature, but also the applied stress, the strain rate, and the temperature of the test, that determine their relative importance. The higher the temperature or the lower the slip resistance, the lower the strain rate, the lower the stress, and the longer, therefore, the duration of the test, the more will the time effect out-

weigh the cycle effect; the smaller, therefore, the difference between the relations expressing fracture stress as a function of time of loading obtained from creep tests and the relations between fracture stress and number of cycles obtained from repeated load tests, if the accumulated duration of all loading cycles to fracture is considered as duration of the load and the

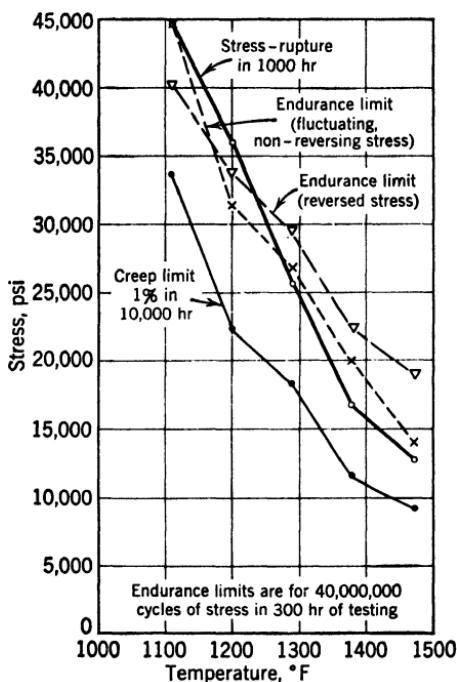


FIG. 58·1b Comparison of creep strength and endurance limits as functions of temperature (Ni-Cr-W-Cb-Fe alloy).

steady (mean) stress as the applied stress. The functions of time versus creep-fracture stress and fatigue stress, respectively, would thus tend to become identical with increasing temperature of the test. This fact is borne out by tests<sup>58·6</sup> (Fig. 58·1). The more, under repeated load cycles, the time effect outweighs the cycle effect, the more does the fatigue strength depend on the frequency of the load cycles, decreasing with decreasing frequency because of the increasing time required for the application of a specified number  $N$  of load cycles.

At very high frequencies the loading rate might attain or even exceed the order of magnitude of the rate of unblocking of slip planes by diffusion of interstitial atoms (see Art. 20); moreover the very short duration of the unloading part of the cycle may prevent the disorientation by rotation along the inter-crystalline regions of the crystal fragments produced during loading. Thus the amount of energy dissipated per cycle will necessarily be much lower at very high frequencies than at the usual frequencies at which fatigue tests are performed, and so will the amount of damage. Beyond a certain limiting frequency the fatigue life of metals, particularly strain-aging metals, will increase and the damping associated with fatigue will decrease.

Fracture of amorphous materials under various conditions is essentially determined by the momentary strain rate; the effect of the deformation preceding it is considerably less important than in metals, unless changes of configuration or anisotropy are produced by such deformation. The lower at the same stress and temperature the strain rate applied, the more extensive the internal redistribution of the response to the external load by relaxation within the structural elements with the shortest relaxation times. With increasing strain rate or decreasing temperature the relation between the duration of the load and the shortest relaxation times, characteristic of the instable elements, may change so that separation becomes almost perfectly brittle since no inelastic deformation can occur. Where the separation strength of a polyphase material is essentially due to the cohesive strength of a continuous network of chain molecules, both an apparent yield limit and an endurance limit will exist delimiting, respectively, a state of extensive disruption of the network and of localized initiation of such a disruption process.

### 59. Condition of Fracture

Since fracture depends essentially on the preceding and the accompanying inelastic deformation, it is reasonable to assume that the phenomenological criterion of fracture also depends on inelastic deformation; it must therefore depend on the principal variables affecting inelastic deformation, that is, stress, strain and their time derivatives, temperature, and previous strain history. In addition, it must depend on the volumetric expansion in a

different way than on distortion, since this expansion influences the separation process, whereas (because of its perfect reversibility) it cannot affect the process of inelastic deformation.

If the initiation of the fracture process within statistically isotropic materials depends essentially on previous deformation, it cannot depend on the direction of the selected coordinate system as long as the deformation is isotropic; the fracture criterion of the isotropic material can therefore be introduced in terms of the invariants of either stress or strain. Since volumetric expansion affects fracture considerably more than volumetric compression, the effect of volume change, which does not interfere with the isotropy of the fracture condition, introduces into it an asymmetry with regard to zero volume change.

Because of the symmetry of the functions defining elastic and inelastic distortion (see Art. 30), the condition of fracture should also be expressed in terms of invariants containing even powers of stress or strain only; because of the asymmetry with regard to zero volume change, it should depend on invariants containing odd power of stress or strain. Since the process of bond separation which initiates fracture depends on the momentary elastic strain or the potential energy in a different manner from what it depends on the inelastic strain or the dissipated energy, the elastic and inelastic strain energies must enter the fracture condition separately. In fact, the elastic and inelastic components of strain influence fracture in the opposite way: Increasing elastic strain increases the tendency to bond disruption, whereas increasing inelastic strain expresses the intensity of the alternative dissipation mechanism, which reduces the tendency to separation.

Neglecting powers of strain of higher than second order, it appears therefore that the second invariant of the deviator of elastic strain (or of stress) in conjunction with the first invariant is the simplest possible measure defining a criterion of fracture. Because fracture is an alternative dissipation mechanism of strain energy, the criterion of fracture can only be an energy criterion, defining a limiting amount of distortional energy stored up prior to fracture in the same way, as the yield limit in terms of resilience defines the conditions of insetting inelastic deformation. The essential difference in the form of these two conditions is the appearance of the first invariant of elastic strain in the condi-

tion of fracture. The dependence of the critical fracture energy on the structure of the material changed by the preceding and the accompanying inelastic deformation can be expressed as a dependence on the dissipated energy  $W_D$  which has been expended in producing the change during the preceding deformation, and on the rate of energy dissipation  $\dot{W}_D$ , associated with the inelastic deformation accompanying fracture.

The simplest condition of fracture can thus be written in the form,

$$\Phi_0(I_{0e2}) = \Phi_0(I_{0s2}) = F[W_D, \dot{W}_D, T, I_{e1}] \quad (59 \cdot 1)$$

where  $c$  denotes the *recoverable* strain and  $F$  is a linear function of the parameters  $W_D$ ,  $\dot{W}_D$ ,  $T$  and  $I_{e1}$ . In order to make eq. 59·1 dimensionally homogeneous the influence of volumetric expansion should be introduced as the square of the first invariant rather than as the invariant itself. The asymmetry of the volume effect with reference to volume-constant deformation would then require the introduction of  $I_{e1}^2$  with such an experimental multiplier which would represent correctly the effect on the fracture strength of a state of volumetric expansion.

Expressing the invariant  $I_{e1}$  by the energy of volumetric expansion  $\Phi_v = KI_{e1}^2$ , multiplied by a linear factor  $c$ , and transferring it to the left-hand side of eq. 59·1, the fracture criterion may be written in the form:

$$\Phi_0 + c\Phi_v = \Phi_s + (c - 1)\Phi_v = F(W_D, \dot{W}_D, T) \quad (59 \cdot 2)$$

Under the simplest assumption  $c = 1$  for  $I_{e1} > 0$  and  $c = 0$  for  $I_{e1} < 0$ , the left-hand side becomes the total potential or elastic strain energy  $\Phi_s = \Phi_0 + \Phi_v$ . For definite values of  $W_D$  and  $\dot{W}_D$  and  $T$ , eq. 59·2 is thus transformed into the fracture condition,

$$\Phi_s = \text{const for } I_{e1} \geq 0 \quad \text{or} \quad e_v \geq 0 \quad \text{and}$$

$$\Phi_0 = \text{const for } e_v < 0 \quad (59 \cdot 3)$$

which was originally proposed by Huber and has been referred to in Art. 22.

There is, however, no justification for the assumption of  $c = 1$  for  $e_v \geq 0$ . On the contrary, it appears reasonable to assume that the effect of the energy of volumetric expansion on the initiation of a process of bond separation is of an intensity different from that of the energy of elastic distortion. Only by

considering  $c$  to be a nonlinear function of  $e_v$  can the asymmetry of the effect of volumetric change on the condition of fracture be adequately expressed, if the function  $c(e_v)$  or  $c(I_{e1})$  is determined by experiment. Hence, the general condition of fracture can be written in the form,

$$\Phi_0 = F_1(W_D, \dot{W}_D, T) - c(e_v)\Phi_v \quad (59.4)$$

which is no longer limited to powers of strain of second order. Whereas  $\Phi_v$  is positive for both volumetric expansion and compression,  $c(e_v)$  is positive for  $e_v > 0$  and negative for  $e_v < 0$ .

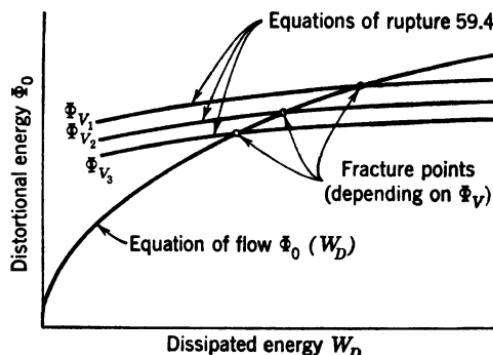


FIG. 59.1 Comparison of equations of flow and equations of fracture.

The comparison of the fracture functions (eq. 59.4) and the flow function  $\Phi_0(W_D, \dot{W}_D, T)$  shows that if fracture is defined by the point of intersection of these functions the extent of deformation at fracture is a function of the volumetric strain. The larger the spread of the fracture curves due to the effect of volumetric strain, the larger the differences in fracture strain. If fracture occurs within a range of deformation in which the rate of work hardening is small, the fracture stress will vary only slightly with the superimposed volumetric strain, whereas the differences of strains will be very large (Fig. 59.1).

For conditions of vanishing volumetric strain the fracture condition and the work-hardening function have the same form  $\Phi_0 = F(W_D, \dot{W}_D, T)$ ; only the absolute value of the critical distortional energy for given values of the parameters are different for fracture and for inelastic deformation. Thus, for instance, both static and dynamic fracture in torsion will necessarily be

governed by a distortion energy condition. This condition will however be inadequate to describe the results of fracture tests under biaxial tension. This conclusion is confirmed by results of both static and fatigue test under combined stresses.<sup>59-1</sup>

The effect of the state of stress on the fracture strength of isotropic materials has not yet been reliably established by experiment. Most of the experimental evidence available has been obtained from notched-bar tension tests, the results of which are difficult to interpret, since the stress distribution in the plastically deformed notch is not known. However, introducing certain approximations concerning this distribution, McAdam has found that the shear stress at which fracture occurs decreases with increasing hydrostatic tension.<sup>59-2</sup> Since the shear stress in a biaxial state of stress is proportional to  $\sqrt{\Phi_0}$  and the hydrostatic tension to  $\sqrt{\Phi_v}$ , this observation would tend to confirm the general form of the fracture condition (eq. 59-4).

The principal difficulties in the experimental procedure and in the interpretation of results of fracture tests under a general state of stress are in the design of specimens in which a measurable state of triaxial stresses can be obtained and which are not initially anisotropic as a result of the manufacturing process (drawing, rolling), as well as in the evaluation of the anisotropy produced in the course of plastic deformation and work hardening.

In tests of tubes under biaxial homogeneous stresses the fracture strength was found to increase when the transversal and longitudinal stress were of the same sign and to decrease when they were of opposite sign. These observations are consistent with the fracture condition (eq. 59-4), since, with increasing biaxiality of the state of stress, the stress components necessary to produce a specified value of distortion<sup>2</sup> energy increase. As the dependence of the fracture limit on  $W_D$ ,  $\dot{W}_D$ , and  $T$  is introduced essentially by the dependence of the deformational behavior on those variables, it can be assumed that, the more important the effect of any one of the variables with regard to deformation, the more pronounced its effect on the fracture strength.

The effect of strain history on fracture stress is mainly the result of the effect of strain history on inelastic deformation. It has been observed that at room temperature compressive pre-strain of moderate extent reduces the fracture strength in ten-

sion.<sup>59-3</sup> This appears to be the consequence of the "Bauschinger" effect produced by the latently stored energy of the textural stresses due to crystal fragmentation under the compressive prestrain. If this latent energy potential, particularly the latent energy of the volumetric expansion responsible for the density decrease associated with work hardening, is introduced into the fracture condition, this condition becomes sensitive to the direction of applied strain. The latent energy produced by the prestrain reduces the amount of strain energy that can be stored up prior to fracture, by intensifying the influence of volume expansion on the limit of fracture.

Hence, if the work hardening is accompanied by hydrostatic pressure which reverses the volumetric expansion due to crystal fragmentation, the limiting amount of energy that can be stored up prior to fracture is necessarily increased. This conclusion has been confirmed by fracture tests after restraining the specimen in tension under various hydrostatic pressures.<sup>59-4</sup> The higher the hydrostatic pressure under which a certain amount of prestrain was produced, the higher the recorded fracture stress and the larger the fracture strain of the specimen.

The fracture strength in tension is however increased by compressive prestrain, if this prestrain is no longer moderate, so that the effect of the work-hardening crystal fragmentation followed by texture formation, that is, the over-all grain refinement and developing anisotropy exceeds the effect of the textural stresses built up within the structure under moderate prestrains. Thus, both the damaging effect of moderate prestrain as well as the strengthening effect of large prestrain are due entirely to the effect on plastic deformation preceding and accompanying fracture, not to any direct effect on fracture strength. Therefore, the fracture strength in tension is invariably increased by tensile prestrain.

The validity of the fracture condition (eq. 59-4) depends on the initial isotropy of the material before, as well as on its isotropy during, the deformation. Since large plastic strains in polycrystalline metals are usually accompanied by textural anisotropy, fracture preceded or accompanied by such strains cannot be governed by a condition based on isotropy of the deformed material. Hence, the simple superimposition of the effects of various states of prestrain by adding the amounts of energy  $W_D$  dissipated in producing the inelastic strain, as sug-

gested by eq. 59·4, cannot be adequate to describe the dependence of fracture strength on strain history, unless the strains are small enough not to affect the isotropy of the material.

The observed slight dependence of the fracture criterion on states of moderate hydrostatic pressure is a property of statistically isotropic and homogeneous materials of very low compressibility only. The more inhomogeneous and compressible a material, the stronger the effect of hydrostatic compression and

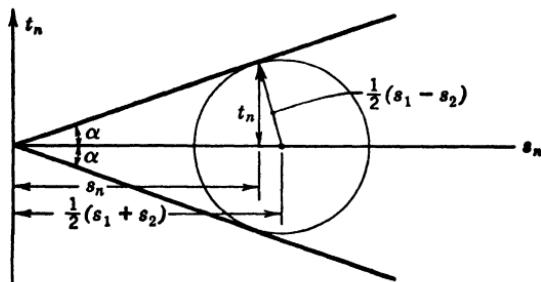


FIG. 59·2 Enveloping lines of Mohr's stress circles for granular material.

tension on the fracture strength. The larger the elements of which the material is built up and the smaller the binding surface forces between the elements, the more pronounced both the strengthening effect of volumetric compression and the weakening effect of volumetric expansion. This increased effect is probably the result of local disruption of cohesion within the inhomogeneous structure under homogeneous expansion, and of the increasing importance, under volumetric compression, of the friction between the particles in providing resistance to deformation.

Fracture tests of concrete, of stone, and of metals of microscopically inhomogeneous structure, such as cast iron, show a very pronounced influence of hydrostatic pressure on fracture stress.<sup>59·5</sup> Although the general fracture criterion for these materials remains an energy criterion, it is convenient to plot the results of fracture tests under biaxial stress in terms of the enveloping curves of Mohr's circles (Fig. 59·2). Since the root of the second stress invariant is very nearly proportional to the principal shear stress  $t_{s\max}$ , the condition,

$$t_{s\max} = f(\sqrt{I_{e1}}) = f(e_v) \quad (59\cdot5)$$

replaces the energy condition.

With vanishing cohesive forces the continuous solid is transformed into a granular material consisting of relatively large discrete particles in which the difference between inelastic deformation and rupture no longer exists, since *gliding* is the only condition of failure. As any continuity of mechanical behavior in such material depends entirely on the existence of a state of hydrostatic compression, the limiting condition (eq. 59·5) is transformed into the well-known linear relation on which the theory of granular materials is based:

$$t_{s_{\max}} = p \sin \alpha \quad \text{or} \quad t_n = s_n \tan \alpha \quad (59 \cdot 6)$$

where  $\alpha$  is the constant slope of the enveloping curve (Fig. 59·2).

## 60. Fracture on Release of Load

Among the most complex phenomena of fracture are those taking place on load *release*; evidently fracture on load release would not be possible without previous large inelastic deformation. On the other hand, the fracture process itself is brittle, since rapid stress release is an elastic process; fracture that follows immediately on load release is therefore not accompanied by inelastic deformation.

There are two different effects which, individually or in combination, lead to such fracture:

1. A state of highly inhomogeneous residual stresses following on large inelastic deformation, the peak intensity of which exceeds the brittle fracture strength of the material.

2. Volumetric expansion on load release, relative to a previous state of large inelastic deformation accompanied by high hydrostatic compression.

In one of the experiments performed by Bridgman<sup>60·1</sup> a thick-walled, hardened steel cylinder was subjected to sufficient external pressure to produce radial plastic yield toward the center of the cylinder, resulting in a permanent decrease of the diameter of the inside hole. On release of pressure, *and after standing for some hours*, a radial crack developed at the inner wall and gradually spread outward, until it reached the outside of the wall. It is also known that glass cylinders, subject to sustained outside pressure, tend to crack spontaneously sometime after release of pressure. Although it might be concluded that the phenomenon is produced by the residual tensile stresses of the

deformed cylinder, two of its aspects cannot be explained by residual stresses alone: the spreading of the crack through the *entire* thickness of the wall, and the delay in starting the crack. There is, however, an explanation for both, if it is considered that under the superimposed hydrostatic pressure the distortional stress produces an irrecoverable readjustment to the external loads by place change of particles within the atomic structure, and that the new configuration of particles has become stable under the applied hydrostatic pressure. The release of the distortional component produces residual stresses; the subsequent release of the hydrostatic pressure acts now like an imposed volumetric expansion with regard to the structure of the material which has readjusted itself to the distortional stresses under the high hydrostatic pressure. Thus, after the instantaneous reversal of elastic strain on load release, a delayed strain recovery sets in, which is an *aftereffect* produced in the metal cylinder by the interaction of the previously deformed intercrystalline regions with the crystal regions which, in addition to their distortion, have undergone a purely elastic compression and therefore attempt to recover their initial volume; in amorphous materials a similar effect is produced by the differences in the relaxation times of the various phases or elements.

The relaxation of instable bonds, followed by a gradual concentration of the response to the system of residual stresses in a decreasing number of bonds, leads to delayed separation on the atomic scale, which is intensified by the relative volumetric expansion. The crack which becomes visible after the atomic separation has become extensive (which takes some time) is the result of the combined effect of residual stresses, relaxation, and relative volumetric expansion. The spreading of the crack through the entire wall thickness, not only through the zone of tensile residual stresses, is due entirely to the relative volumetric expansion.

Zener has illustrated the effect of relaxation by his discussion of the observation that the tip of a projectile, recovered after passing through an armor plate, has a tendency to fly off spontaneously after a certain time.<sup>60-2</sup> Presumably the plastic deformation of the tip during the passage of the projectile through the armor has taken place under substantial volume compression; the residual stresses are therefore accompanied by a volumetric

expansion relative to the new stable configuration produced during the deformation under simultaneous compression. Relaxation within the intercrystalline boundaries gradually shifts the resistance to the residual stresses into a decreasing number of bonds, a process which, after some time, is bound to produce separation along the plane normal to which the relative extension is largest. This is the plane normal to the direction of the highest compression, which is located close to the boundary between elastic and plastic deformation. It is along this plane that the tip flies off. Placing the projectile in boiling water intensifies the relaxation and therefore shortens the interval after which the tip flies off; low temperatures, on the other hand, by reducing relaxation, cause delay of the phenomenon or prevent its occurrence altogether.

Another illustrative example of fracture on load release has been reported by Griggs.<sup>60-1</sup> A specimen of limestone was subjected to a uniaxial compressive stress under a superimposed high hydrostatic pressure. The uniaxial compressive stress was released first after considerable inelastic deformation had occurred; when the hydrostatic pressure was subsequently released, the specimen ruptured into thin disks along planes perpendicular to the direction of the uniaxial compression. If the permanently deformed configuration is considered as the zero point, the release of the compressive load acts as uniaxial tension, whereas the release of pressure acts as hydrostatic tension. At a certain combination of both, fracture occurs in the direction of maximum extension relative to the permanently deformed state.

The previous example illustrates the formation of shale out of an isotropic layer. This layer is first compressed and considerably deformed by a transient directional force under the high hydrostatic pressure which acts at the depth at which the layer is situated. When subsequently, and long after the transient force has vanished, the layer is brought up nearer to the surface with resulting total or partial release of the hydrostatic pressure, it develops the characteristic separation surfaces normal to the direction of the vanished force.

The condition of fracture on load release can phenomenologically be formulated in terms of the residual stresses and of a relative volumetric expansion, the amount of which should be

calculated from a new zero point, defined by the stable configuration reached under the applied external forces. The length of the period of application of these forces and the extent of the resulting inelastic deformation would therefore appear to be of primary importance. Since distortional stresses acting for short times generally do not produce extensive permanent readjustment of the configuration within the material, phenomena of fracture on release of loads of short duration would be time-independent, since they can be due to the residual stresses alone.

### References

- 55·1 J. H. HOLLOMON and C. ZENER, *J. Applied Phys.* **17** (1946) 78.  
J. H. HOLLOMON, *Welding J.* **25** (1946) 575.
- 55·2 A. GRIFFITH, *Phil. Trans. Roy. Soc. A* **221** (1920) 163.
- 55·3 C. E. INGLIS, *Inst. Naval Arch., London* **60** (1913) 219.
- 55·4 J. I. FRENKEL and T. A. KANTOROVÁ, *J. Phys. USSR* **7** (1943) 108.
- 55·5 W. WEIBULL, *Ing. Vetenskap Akad. Handl. Stockholm* **151-53** (1939).
- 55·6 R. A. FISHER and L. H. C. TIPPETT, *Proc. Cambridge Phil. Soc.* **24** (1928) 180.
- 55·7 H. CRAMER, *Mathematical Methods of Statistics*, Princeton Univ. Press (1946) 373.
- 55·8 B. EPSTEIN, *J. Am. Statist. Assoc.* **43** (1948) 403; *J. Applied Phys.* **19** (1948) 140.
- 55·9 A. M. FREUDENTHAL, *Proc. Roy. Soc. A* **187** (1946) 416.
- 55·10 H. E. DANIELS, *Proc. Roy. Soc. A* **183** (1945) 405.
- 55·11 R. FUERTH, *Proc. Roy. Soc. A* **177** (1940-41) 217.
- 58·1 J. H. HOLLOMON and C. ZENER, *Trans. AIME* **158** (1944) 283.
- 58·2 C. S. BARRETT, *The Structure of Metals*, McGraw-Hill Book Co., New York (1944).
- 58·3 D. J. McADAM and R. W. MEBS, *Proc. ASTM* **43** (1943) 661.
- 58·4 N. F. MOTT, *Proc. Roy. Soc. A* **189** (1947) 300.
- 58·5 A. M. FREUDENTHAL and T. J. DOLAN, *4th Progress Rept. on Fatigue*, Office Naval Research and Univ. Illinois Eng. Expt. Sta., Urbana.
- 58·6 R. B. SMITH, *J. Applied Mechanics* **14** (1947) A-101.
- 59·1 H. MAJORS, B. D. MILLS, and C. W. McGREGOR, *J. Applied Mechanics* **16** (1949) A-7.
- 59·2 D. J. McADAM, *Metals Technol.* (1944) T.P. 1872.  
D. J. McADAM, G. W. GEIL, and R. W. MEBS, *Proc. ASTM* **45** (1945) 448.
- 59·3 D. J. McADAM, G. W. GEIL, and W. H. JENKINS, *Proc. ASTM* **47** (1947).
- 59·4 P. W. BRIDGMAN, *J. Applied Phys.* **17** (1946) 201.
- 59·5 TH. V. KÁRMÁN, *Z. Ver. deut. Ing.* **55** (1911) 1749.
- 60·1 P. W. BRIDGMAN, *op. cit.*, **9** (1938) 517.
- 60·2 C. ZENER, *Metals Technol.* (1946) T.P. 1992.

## CHAPTER

## 12

RHEOLOGICAL BEHAVIOR OF SUSPENSIONS  
AND GELS**61. Brownian Motion. Thixotropy**

Suspensions and gels are forms of aggregation of matter that are intermediate between fluids and solids. They are of great importance in many processes of transition from the fluid to the solid state, apart from having intrinsic industrial significance. Both suspensions and gels are fluid-containing systems of various composition, with solid particles in different degrees of dispersion. Whereas suspensions are essentially liquids, gels can be considered as a form of solid since they possess a yield limit and therefore elasticity and rigidity below this limit.

According to the degree of dispersion of the solid particles, suspensions are divided into molecular, colloidal, and coarse suspensions. Since both the intensity of thermal activation and the magnitude of interacting forces depend essentially on particle size, the size of the solid particles is an important characteristic. The range of particle sizes has been presented in Fig. 8·1.

Molecular suspensions are the chemical solutions that behave like true liquids. Their particle size is of an order of magnitude  $< 10^{-7}$  cm. Colloidal suspensions, also called *sols*, contain solid particles of colloidal size (between  $10^{-7}$  and  $10^{-5}$  cm); although they cannot be made visible in their individual shape unless they are near maximum size, they are not dissolved. The particles are small enough not to settle under the forces of gravity, but also large enough to have only a relatively slow rate of diffu-

sion. If the size of the dispersed particles is appreciably larger than  $10^{-5}$  cm, the suspension is called coarse. Whereas molecular suspensions are clear, and colloidal suspensions usually appear clear unless they are strongly illuminated and, by reflecting the light, produce a certain opacity, coarse suspensions are visibly turbid.

Particles in coarse dispersion can in general be separated from the fluid medium by gravitation, centrifugal forces, and mechanical filters, unless the particles in the apparently coarse suspension consist of or are surrounded by loosely bound groups of colloidal particles. Dispersed colloidal particles remain in suspension. The dispersed phase of a coarse suspension, however, tends to settle either upward or downward, according to the relative densities of fluid medium and particles. The settling or *sedimentation* will evidently be the less rapid, the higher the viscosity of the medium. The separation of the solid particles of a coarse suspension from the medium may be accelerated by centrifugal forces, whereas such forces exert no influence on true solutions or colloidal suspensions.

When the behavior of a colloidal suspension is considered, the question arises why solid particles of a density that is higher than the density of the medium in which they are suspended remain in suspension instead of settling down. The answer to this question is found in the existence of the *Brownian motion* of small solid particles in a gaseous or liquid medium. This motion is a phenomenon associated with the size of the particles, and its discovery and explanation was one of the most important steppingstones in the development of the molecular theory of matter.

The Brownian motion is a random motion which can be observed for all suspended particles of an order of magnitude not exceeding  $10^{-5}$  cm. Its intensity varies inversely as the size of the particles;  $10^{-4}$  cm appears to be the maximum limiting diameter of particles above which no motion can be observed. The motion overcomes the forces of gravitation and tends to distribute the particles throughout the fluid with statistical uniformity. It is due to the collision of the suspended particles with the thermally agitated molecules of the liquid and thus provides visible evidence of the molecular theory of matter. The intensity of the motion is increased by light and by radiant

as well as by conducted heat and is also affected by the electric charge of the particles and by different chemical reagents by which the motion is either accelerated or retarded.

A molecular (statistical) theory of the Brownian motion has been developed independently by Einstein and Smoluchowski.<sup>61-1</sup>

*Gels* are formed from colloidal suspensions or sols by increasing the concentration of the dispersed solid phase either by addition of solid substance or by partial evaporation of the fluid medium, by decrease of temperature or by chemical reactions (polymerization, flocculation). The common feature of all those processes is the gradual immobilization of the liquid phase by the formation of a structure within the dispersed phase; the gel structure depends on the manner in which the immobilization has taken place.

In general the structure of gels is assumed to consist of coherent networks of adherent particles, filled with fluid, both phases of the solid-liquid substance being continuous. A variation of this structure is a form in which the solid component only is continuous, enclosing the liquid in the form of isolated drops. X-ray diffraction patterns of some gels have revealed interference rings; it appears therefore that a certain degree of order is introduced by the formation of the network. Because of the continuity of the formed network, flow cannot proceed so long as this network remains intact under the applied shear force. The smallest value of the shear stress necessary to destroy the gel structure sufficiently for flow of the immobilized fluid medium to occur represents the *yield limit* of the material.

The appearance of a yield limit, that is, the formation of a gel structure, must not necessarily be visualized as the simple effect of mechanical contact of particles. Experiments have shown that a distinct yield limit can be observed in suspensions of concentration of about 15 to 25 percent, and even at 2 percent (volume).<sup>61-2</sup> For such concentrations it can only be the forces of interaction between the suspended particles in their equilibrium positions that resist the applied shear until the force system is disrupted and the equilibrium arrangement of particles destroyed. The existence of a yield limit could therefore not depend on such a minimum volume concentration of solid particles in the suspension which would produce actual contact between the particles.

Many gels, particularly clays and paints, revert to sols on mechanical manipulation, such as gentle shaking, stirring, or ultrasonic vibration; when left undisturbed they solidify again. This reversible sol-gel transformation is called *thixotropy*; (from the Greek *thixis* = to touch, and *trepo* = change); it is a process that can be repeated indefinitely with the same result.

According to Freundlich and others, thixotropy is a property of those gels in which the packing of the solid particles is very loose and encloses a large amount of fluid.<sup>61-3</sup> It is due to the intensity of the forces of interaction between the suspended particles, which are broken up by the applied shear forces and re-establish themselves slowly on removal of the force if not disturbed by other external effects. The amount of fluid must be sufficient to allow the particles to maintain Brownian motion after the breakup of the structure, as, otherwise, neither could the substance be fluid after being stirred or shaken, nor could the particles come to rest after spontaneously rearranging themselves in their position of minimum potential energy. Hence, for thixotropy to exist, an upper limit of particle size must not be exceeded, which is the same diameter of about  $10^{-4}$  cm that delimits the existence of Brownian motion. When particles are nearly spherical, thixotropic effects can be expected only if their volume concentration is small, since a continuous and rather closely packed structure cannot be broken up to such an extent that Brownian motion becomes possible. Thixotropy will therefore be favored by nonspherical (elongated) shape of particles since nonspherical particles can form a continuous structure at lower concentration than spherical particles, and this structure can more easily be broken up. In general, thixotropy requires a type of structure that is *metastable* in its nature: the forces tending to combine the dispersed particles into a more or less continuous structure, as well as the concentration of particles, must not be so great as to prevent their dispersion when mechanically agitated or to cause rapid reformation of the structure during the agitation or immediately after the agitation has stopped.

There is evidently a very close interrelation between the existence of a yield limit and thixotropy, although the yield limit is defined in terms of a breakdown stress of the structure, whereas thixotropy may be a function of the shear stress, or of the shear

rate, or of the extent of flow, or of a combination of all those variables.

## 62. Viscosity and Yield Limit

The mechanical behavior of suspensions and gels is described by the shear-stress versus flow rate (consistency) diagram (Fig. 62.1), in a similar way to that in which the mechanical behavior

of solids or of apparently solid materials is described by the stress-strain diagram. Deviations of the observed flow diagram of the material from the linear diagram of the Newtonian liquid can be interpreted in terms of structural changes within the material.

Whereas in the study of inelastic materials it is assumed that all considered states are states of equilibrium, the analysis of suspensions and gels is usually based on the assumption of their being in a steady state of flow. The two

mechanical constants appearing in the relation connecting the dynamical and kinematical variables which, for incompressible flow, are the shear stress  $s_t$  and the rate of shear strain or rate of flow  $\dot{g}$  (in fluid mechanics called the velocity gradient  $D$ ), respectively, are the coefficient of viscosity  $\eta$  and the yield limit in shear  $s_{t0}$ .

Suspensions are described by the equation,

$$\dot{g} = f(s_t) \quad (62.1)$$

since liquids are defined by  $s_{t0} = 0$ . Equation 62.1 must therefore fulfill the boundary condition  $\dot{g} = 0$  for  $s_t = 0$ . The general equation of state for gels has the form: ( $s_t > s_{t0}$ )

$$\dot{g} = f(s_t - s_{t0}) \quad (62.2)$$

Assuming linear behavior, eq. 62.1 becomes the definition of the Newtonian liquid,

$$\dot{g} = \frac{1}{\eta} s_t = \phi s_t \quad (62.3)$$

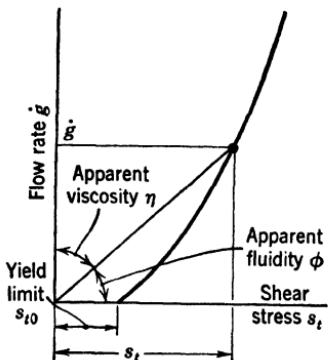


FIG. 62.1 Consistency diagram of gel.

where  $\phi$  is the *fluidity* of the substance, while eq. (62·2) becomes the definition of the Bingham body or viscoplastic body in pure shear flow,

$$\dot{g} = \frac{1}{\eta} (s_t - s_{t0}) = \phi(s_t - s_{t0}) \quad (62 \cdot 4)$$

With the velocity  $\dot{g} \rightarrow 0$ , eq. 62·4 becomes  $s_t = s_{t0}$ , which is the definition of the St. Venant body.

The difference between the concept of the Bingham body as derived here from the viscous liquid and that derived in Art. 42 from the plastic solid is in the relative magnitude of the elastic and the inelastic deformation. Whereas in the mechanics of gels the elastic deformation is considered negligibly small in comparison with flow, so that the material appears perfectly rigid for  $s_t < s_{t0}$ , the analysis of the viscoplastic solid by means of the tensor of overstress considers both elastic and viscoplastic deformation.

A true fluid is characterized by the complete relaxation it exhibits under a stress system different from a pure hydrostatic pressure. This relaxation acts so fast that elastic phenomena practically vanish before they can be observed. Theoretically a certain degree of transient elasticity of shape is present in substances that behave essentially like fluids; it can however be detected only by experiments that are so rapid that they do not provide sufficient time for flow to take place. The higher the viscosity of a liquid, the more relative importance must, however, be attributed to the elastic aspect of its response to external forces. Since both viscosity and rigidity are embodied in the relaxation time, it is in terms of the spectrum of relaxation times that the transition from the perfect liquid with zero relaxation time to the perfectly elastic solid with infinite relaxation time can be described.

The mechanism of flow and of viscous resistance of the true fluid is closely connected with the continuous process of formation and breakup of molecular groups (see Art. 11). The energy dissipated into heat in the course of this statistically uniform structural breakup and re-formation appears as the resistance to flow, expressed by the coefficient of viscosity. Within a liquid with a statistically homogeneous structure there is no reason to expect the specific resistance to flow and therefore the

coefficient of viscosity to depend on the flow rate or on the shear stress, as long as the structure of the liquid is not affected by either flow rate or shear stress. However since suspensions are not true liquids, their behavior cannot be described by eq. 62·3;

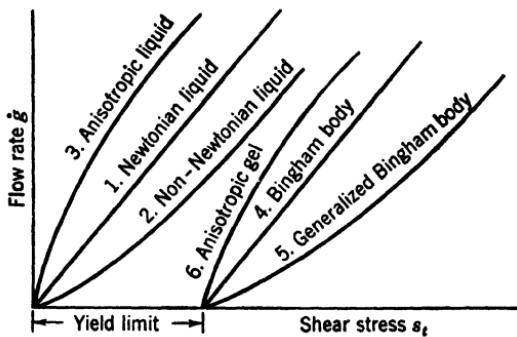


FIG. 62·2 Typical shapes of consistency curves of suspensions and gels.

their viscosity is not constant unless the volume concentration of dispersed particles is very small and the particles nearly spherical and of uniform size.

Gels will usually not show a linear relation after the shear stress has exceeded the yield limit and the substance has become a suspension, for the same reason that suspensions do not show linear viscosity. Considering Fig. 62·2, the behavior of suspensions may follow any one of the curves 1, 2, 3, whereas that of gels follows any one of the curves 4, 5, 6. For suspensions curves with inflection points (Fig. 62·3) have also been observed.

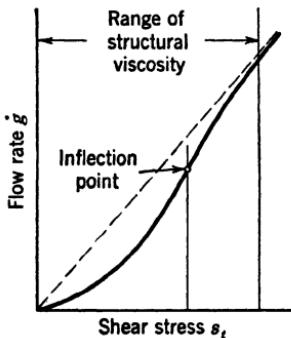


FIG. 62·3 Consistency curve of a suspension with structural viscosity.

a clearly anisotropic structure, such as fibrous materials, and filaments. Normal flow phenomena of highly deformable statistically isotropic substances are usually governed by curves of the types 1, 2, 4, 5, or by the complex curve shown in Fig. 62·3.

Flow relations described by curves 3 and 6, with viscosity increasing with stress, are rather infrequent and restricted entirely to substances with

In considering curves 2, 4, and 5 the question presents itself whether the difference between them is real, that is, whether in most of the deformable substances a yield limit really exists. There is no doubt that the yield limit is considerably less marked in gels than it is in crystalline materials. This is due to the fact that for an apparent yield limit to be observable in gels a *continuous system of interacting forces* is required rather than a *continuous skeleton of solid particles*. Hence, the difference between the configuration of the structure for which a continuous skeleton of particles still exists and for which it no longer exists is not necessarily equivalent to the difference with respect to the existence or nonexistence of a yield limit. Even after the destruction of the mechanical contact of particles, the system of interacting forces may still remain continuous, until the velocity of flow has become large enough to destroy entirely both the initial structure and the system of forces between the particles. Only after this destruction has been achieved does the flow process become truly viscous so that the subsequent formation and breakdown of intermolecular force systems may be considered a purely statistical phenomenon, which can be described by a single constant (the coefficient of viscosity), depending only on the energy expended in the process.

For an undisturbed skeleton or force system a definite yield limit exists at rest; after sufficient agitation this limit may disappear completely. Between the two stages the flow resistance is affected by the degree of destruction of the original skeleton or force system, and the viscosity depends therefore on stress or on flow rate; it becomes practically constant only after a complete breakdown of the structure. It can thus be concluded that the existence of a yield limit is real only for gels with sufficient volume concentration of dispersed particles to form a continuous skeleton, or with sufficiently small particles to produce strong forces of interaction.

The yield limit is identical with the theoretical limit  $s_{t0}$  of the linear Bingham body only where it is due to a continuous skeleton of solid particles of such size that the forces of interaction are relatively small. For gels of different structure the real yield limit is smaller than the theoretical limit of the linear body, since flow in the fluid phase may start by slow change of the equilibrium distance between the suspended par-

ticles, without breakdown of the system of forces connecting them, and proceed slowly, accompanied by a *gradual* weakening and destruction of the force system. The theoretical yield limit, that is, the intercept of the asymptote to the consistency curve on the stress axis represents therefore in this case a fictitious limiting stress, which might have been observed if the destruction of the continuous system of forces of interaction were sudden and not gradual (Fig. 62·4).

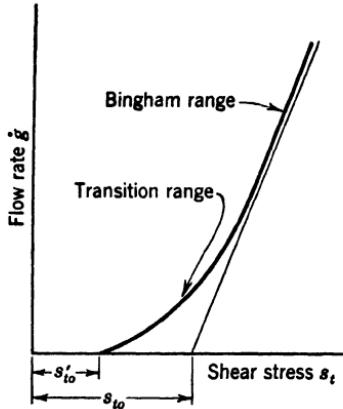


FIG. 62·4 Consistency curve of a gel with real and ideal (Bingham) yield limit,  $s'_{t0}$  and  $s_{t0}$ , respectively.

Curves of type 2 can thus be interpreted in terms of the behavior of Bingham bodies with a fictitious yield limit, whereas curves of type 5 describe substances for which the mechanical contact of solid particles and the field of interacting forces produce phenomena of a comparable order of magnitude. In substances the yield limit of which does not depend on the existence of a mechanically continuous solid skeleton, but only on interacting forces between particles, thixotropy is therefore a perfectly normal phenomenon, as in such

substances the system of interacting forces will always re-establish itself after external agitation ceases and thus re-establish the yield limit. Thixotropy would be an anomalous phenomenon only in substances in which the yield limit is the expression of a mechanically continuous solid skeleton which must not necessarily re-form itself after agitation. If in that case thixotropy comes into play, re-establishing a yield limit, this limit is not necessarily the initial yield limit, but may be a new lower limit, produced by the system of interacting forces alone.

Complex flow diagrams of the type represented in Fig. 62·3 cannot be interpreted in terms of a yield limit since, after a range of changing viscosity, the diagram reverts to a straight line through the origin which describes linear viscous flow. The curved part apparently describes the change of viscosity resulting from the gradual breakup of a system of interacting forces

between nonspherical particles of relatively low volume concentration; attaining random orientation after the complete breakup they are subsequently oriented with increasing rate of flow. It is probably the superposition of the breakup and the orientation effect and the gradual transition from one to the other that produces the inflection point. Because of the easier formation of a continuous structure by elongated particles and because of the possibility of their orientation, the S-shaped diagram is the less probable, the more spherical the shape of the particles, the smaller their volume concentration but also the smaller the interacting forces. On the other hand, a too large volume concentration may produce a yield limit.

The actual behavior of gels is very much affected by the relative magnitudes of yield limit and coefficient of viscosity. There are two principal groups, the differences of behavior of which result from the fact that one group is characterized by a small yield limit and a high coefficient of viscosity, whereas the other group shows a relative high yield limit and a small coefficient of viscosity. The difference in the coefficients of viscosity between the two groups may be as large as  $10^5$ – $10^7$ . Glass, resins and asphalts are prototypes of the first group, whereas clays and paints form the second group. The practical importance of this difference in behavior is considerable: because of their low viscosity, materials belonging to the second group may easily be formed into any shape; the high yield limit, however, ensures that a shape, once given, is retained indefinitely. On the other hand, the high viscosity of glass, resins and asphalts requires the application of large forces if deformation is to be produced, whereas, as a consequence of a relatively low yield limit, the material is bound to lose the given shape in time, even under the influence of gravity.

The reason for the different behavior of materials in these two groups is not only the low coefficient of viscosity of the medium of dispersion of materials belonging to the second group (water, linseed oil) compared to the high viscosity of the media of the first group, but also the difference between the high intermolecular forces within materials of the first group compared to the relatively small forces between the dispersed particles in the second group. Whereas in the first group flow under applied stress of any magnitude takes place by gradual re-formation of

the intermolecular force system, resulting from diffusion and place change of particles, the solid skeleton or the system of interacting forces in substances of the second group is permanently disrupted by the action of applied stress above a certain limit only. Since no re-formation takes place during flow, the minimum breakdown stress represents a true transition point of rheological behavior. It is characteristic that in materials belonging to the first group the medium and the dispersed phase are chemically identical (*isogels*), particles being of molecular size, whereas in the second group they are chemically different (*heterogels*), particles being of colloidal size. Actually gels of the first group form the transition to materials which are usually defined as amorphous solids, although they should, more consistently, be defined as liquids. This uncertainty of definition is an indication of the gradual transition between the different states of aggregation of matter.

### 63. Rheological Measurement. Behavior of Technical Materials

Rheological measurements are concerned with the determination of the relation between stress and rate of flow. For true liquids this relation is expressed by a single coefficient of viscosity. In suspensions and gels the apparent viscosity varies with either stress or flow rate; therefore for such materials the observation of the consistency curve over the whole range of stationary flow replaces the measurement of a single coefficient of viscosity. This is, incidentally, one of the principal practical difficulties in introducing modern rheological methods into industrial laboratories where most viscous substances tested are non-linear; the time required to produce a consistency curve is necessarily a multiple of that required to determine a constant viscosity coefficient.

Conditions of stationary flow are produced by subjecting the material to a simple type of stress, preferably to a homogeneous shearing stress. For every value of stress  $s_t$  a definite steady state of flow of velocity  $\dot{g}$  is reached after a certain period of transient phenomena. The consistency curve of the substance  $\dot{g}(s_t)$  may then be plotted. If the material has an observable yield limit, no flow will occur if the shear stress applied does not exceed this limit. If the experimental conditions set up are

such that the state of stress is nonhomogeneous, the test must be analyzed and the relation with the homogeneous stress established by integrating the hydrodynamic equations for the unit volume under the particular testing conditions. The existence of a distinct yield limit and of linear flow in homogeneous shear does no more imply that a linear stress-flow-rate relation will be observed in a conventional viscometer test, than the existence of a sharp yield limit followed by linear work hardening in the tension test of a metal implies that the same linearity between moment and deformation will be observed in a bending test. Actually, most of the industrial viscometers produce nonhomogeneous conditions of stress, and the relation between the observed mechanical variables must be reinterpreted in terms of the volume element, before the physical meaning of the test becomes clear.

The principal types of viscometers used for suspensions, and of plastometers used for gels are:

1. Capillary-tube viscometers, in which the substance is driven by a pressure  $p$  through a long narrow tube or radius  $r$  and length  $l$ , and the viscosity  $\eta$  determined from the simple so-called Poiseuille-Hagen linear viscometer equation,<sup>63·1</sup>

$$V = \frac{p\pi r^4}{8l\eta} = \text{const } p/\eta \quad (63 \cdot 1)$$

where  $V$  denotes the volume of substance extruded per second. For materials with a yield limit  $s_t \neq 0$  a more complex nonlinear plastometer equation has been derived by Buckingham and Reiner;<sup>63·2</sup> it is transformed into eq. 63·1 for  $s_{t0} = 0$ .

2. Rotation or concentric-cylinder viscometers, in which the substance is placed in the space between two cylinders, the inner one of which is rotated. The material rotates in concentric layers; its viscosity can be determined from the linear relation between the torque  $M$  necessary to produce a definite constant angular velocity  $\omega$  of the inner surface,

$$\eta = \text{const } \frac{M}{\omega} \quad (63 \cdot 2)$$

For materials with a yield limit,  $s_t \neq 0$ , Reiner has derived the appropriate plastometer equation,<sup>63·3</sup>

$$\eta = C_1 \frac{M}{\omega} - C_2 \frac{s_{t_0}}{\omega} \quad (63 \cdot 3)$$

where  $C_1$  and  $C_2$  are instrument constants.

3. Compression visco- and plastometers, in which cylinders of the viscoplastic material are compressed between parallel plates under constant load; the rate of decrease in height or the decrease in height after a time interval is measured as function of the load. The principal difficulties in using parallel-plane plastometers are that the expanding cross section produces changes of stress during the test which must be compensated and that the frictional restraint over the ends of the cylinder cannot be reliably evaluated. This makes the derivation of rational instrument equations very difficult. Various formulas are used, most of them derived under such simplifying assumptions that their validity is very doubtful.<sup>63·4</sup>

4. Usual mechanical tests of viscoelastic materials, such as torsion or tension tests of cylinders, or bending tests under constant moment or under the own weight of the specimen (sagging beam test). The viscosity coefficient is obtained by using the elastic-viscoelastic analogy in the interpretation of observations, adapting the known elastic solution of the problem by replacing  $g$  by  $\dot{g}$  and  $G$  by  $\eta$ .<sup>63·5</sup>

5. Purely empirical or comparative tests, such as penetration or indentation hardness tests, and other specific tests for various materials; these tests do not give true physical constants and can only be interpreted on the basis of empirical formulas.

Viscometer and plastometer tests are very important in the manufacture of a large variety of materials, the technology of which forms a series of processes dominated by rheological behavior. Among the rheological materials of primary engineering importance are cement, asphalt, natural and synthetic resins, paints and clays. The mechanical behavior of rubber-like materials and of yarns is frequently determined by conventional mechanical tests, observing stress and strain instead of stress and flow rate.

The rheological properties of asphalts vary widely. Values of viscosity at room temperatures range from about  $10^3$  poises for the fluid petroleum residues to between  $10^{10}$  and  $10^{12}$  for hard asphalts. The type of flow varies from that of an essentially

Newtonian liquid to a flow of high complexity. Because of their complex nature, asphalts cannot be described on the basis of their chemical composition alone, and it is necessary to classify them by physical tests.

Asphalts show definite elasticity and therefore both aftereffect (creep recovery) and stress relaxation. They also show age hardening, that is, an increasing value of the apparent viscosity with time, and thixotropy. As may be expected, the presence of thixotropic structure and age hardening is closely interrelated with the existence of a yield limit in the asphalts of continuous skeleton structure. Relaxation times near room temperatures are of the order of magnitude of a few seconds.<sup>63-6</sup>

The conventional flow tests of asphalt are purely empirical; the three most frequently used conventional tests are the penetration test (ASTM-D5-25), the "ring and ball softening-point" test (ASTM-D36-26) and the ductility test (ASTM-D113-39). These tests, very useful for obtaining comparative data, are inadequate to indicate fundamental properties since they measure combinations of basic properties. The flow under penetration is very much different from that in a viscometer; there is, moreover, a conical depression around the needle due to surface effects and effects of adhesion of steel and asphalt. The "softening-point" test measures a combination of viscosity, density, thermal conductivity, all of which vary with temperature and with the complexity of the flow. The ductility, because of the constant deformation rate imposed, is the expression of the effect of a rapidly increasing tensile stress under which the cohesive strength of the material is reached. Among the three conventional tests the ductility test is, however, the only one that can be interpreted in terms of basic mechanical properties.

Because of the relatively high strength and high viscosity, mechanical behavior and properties of resins are usually observed in conventional mechanical and not in rheological tests. However, the flow properties of resinous materials are important during the different phases of their manufacture when their structure has not yet been sufficiently built up to produce essentially solid behavior. An example of such an application is the rheological testing of crude rubber. In this case the observed flow properties have not necessarily a meaning of their own but may be used as an indication of certain properties required which,

before the introduction of the rheological testing, may have been inferred from a visual inspection or from a so-called *psycho-physical test*.<sup>63-7</sup>

The apparatuses used for rubber and rubber compounds are either the parallel-plane plastometer, the extrusion viscometer, or the rotating-cylinder viscometer. The parallel-plane plastometer has actually been developed for rubber testing. The variable that is usually observed in the test is the thickness of the sample after a certain period of compression under a standard weight and at a standard temperature; in recovery measurements the observed characteristic is the recovery in height on release of load. Even if the developed theoretical formulas for calculating viscosities from such measurements could be assumed to be fairly accurate, they cannot be applied to raw rubber, because of the considerable effect of thixotropy. Thixotropy interferes also with the operation of the extrusion viscometers. The most appropriate type for viscosity measurement of rubber is the rotating-cylinder viscometer in which both viscosity and elastic recovery can be observed.

Paints are very nearly ideal Bingham bodies with viscosity of the order of 10 poises and with yield limits of the order of magnitude of  $10^2$  dynes per  $\text{cm}^2$ . The viscosity makes the paint *brushable*; the yield limit prevents it from sagging or running off, but if it forms too rapidly (thixotropy) it prevents leveling of the brush marks; thus the concentration of pigments and the viscosity of the oil must be selected to ensure the best performance in all respects.<sup>63-8</sup>

In clays the fluid medium is water; the particles have the shape of rough leaves or flakes. The plasticity of clays varies with the proportion of water; as water is added to dry clay, its plasticity increases until a maximum is reached. A further addition of water reduces the plasticity, so that it can no longer keep a given shape and becomes first sticky, then a fluid slurry. The plastic properties of clays are related to the flakelike shape of the particles, to the roughness of their surface, to the electrostatic forces between them, and to the water films forming around them. Their rheological behavior approaches that of a Bingham body.<sup>63-9</sup>

Thixotropic effects are considerable; it is therefore to be expected that the flow of clay should be non-Newtonian.

Although the linearity of the flow diagram for higher stresses indicates the existence of a definite theoretical yield limit  $s_{t_0}$ , flow starts actually at a lower limit  $s_{t_0}'$ . Terzaghi<sup>63-10</sup> has called  $s_{t_0}'$  the flow limit and  $s_{t_0}$  the yield limit, defining the former as a limit at which viscous flow starts within the layers of water without changing the structure of the particles, and the latter as a limit at which the water film is disrupted and the bonds between adjoining particles are broken.

Thixotropy of clays is of considerable importance in the interpretation of results of tests in soil mechanics. Deep layers of clay deform under the weight of the upper layers. This deformation proceeds by *consolidation*, that is, by the squeezing out of water from the pores of the clay layer which, therefore gradually decreases in volume and, consequently, in depth. If thixotropic effects are present, the consolidation which, according to the theory, proceeds asymptotically until an equilibrium is reached between the hydrostatic pressure within the clay pores and the imposed pressure,<sup>63-11</sup> may be stopped by the development of a continuous system of interacting forces, the resistance of which is sufficient to carry the overload, so that a hydrostatic gradient no longer exists within the water surrounding the clay particles. Thixotropic effects also interfere with the taking of clay samples for soil investigations, since the process of drilling out of the core, however carefully performed, is bound to disturb the structure of internal forces which has been formed over the long periods during which the clay deposit has been at rest. So far, no procedure has been found by which it would be possible to re-establish within short periods in the laboratory the resisting structure of the natural-clay formation which has developed in the course of long periods of consolidation.

#### References

- 61·1 A. EINSTEIN, *Ann. Physik* (4)17 (1905) 549; 19 (1906) 371.  
M. v. SMOLUCHOWSKI, *Ann. Physik* (4)21 (1906) 756.
- 61·2 *2d Report on Viscosity & Plasticity*, Acad. Sci. Amsterdam (1938) 229.
- 61·3 H. FREUNDLICH, *Thixotropy*, Herman & Cie., Paris (1935).
- H. B. WEISER, *Colloid Chemistry*, John Wiley & Sons, New York (1939).
- 63·1 M. REINER, *Ten Lectures on Theoretical Rheology*, R. Mass, Jerusalem (1943) 71.
- 63·2 E. BUCKINGHAM, *Proc. ASTM* 21 (1921) 1154.  
M. REINER, *Kolloid-Z.* 39 (1926) 80.

63·3 M. REINER and R. RIVLIN, *Kolloid-Z.* **43** (1927) 72.  
63·4 *Op. cit.*, 61.2, 221.  
63·5 M. REINER, *op. cit.*, 66.  
63·6 R. HOUWINK, *Elasticity, Plasticity, and Structure of Matter*, Cambridge Univ. Press (1937) 165.  
63·7 G. W. SCOTT BLAIR, *A Survey of General and Applied Rheology*, Pitman, London (1944) 154.  
63·8 R. HOUWINK, *op. cit.*, 311.  
63·9 *Ibid.*, 345.  
63·10 K. TERZAGHI, *J. Rheol.* **2** (1931) 253.  
63·11 K. TERZAGHI, *Theoretical Soil Mechanics*, John Wiley & Sons (1943) 265.

P A R T  
C

## Application of the Mechanics of Inelastic Behavior



## PLASTICITY. PROBLEMS OF EQUILIBRIUM

### 64. Torsion

The simplest equilibrium problem of the ideal elastic-plastic body is that of torsion. In the case of elastic torsion the problem is formulated by assuming that the stresses are independent of the direction  $x_3$  of the axis of the member so that all differentials with regard to  $x_3$  vanish. Of the equilibrium equations only the equation,

$$\frac{\partial s_{13}}{\partial x_1} + \frac{\partial s_{23}}{\partial x_2} = 0 \quad (64 \cdot 1)$$

remains, since in pure torsion  $s_{11}$  and  $s_{22} = s_{12} = 0$ . Introducing  $t_1 = s_{13}$  and  $t_2 = s_{23}$ , this equation may be written in the form:

$$\frac{\partial t_1}{\partial x_1} = - \frac{\partial t_2}{\partial x_2} \quad (64 \cdot 2)$$

By introducing the function  $\phi$ , so defined that

$$t_1 = \frac{\partial \phi}{\partial x_2}; \quad t_2 = - \frac{\partial \phi}{\partial x_1} \quad (64 \cdot 3)$$

and considering the fundamental elastic equations,<sup>64.1</sup>

$$\Delta s_{23} = \Delta s_{13} = 0 \quad (64 \cdot 4)$$

where  $\Delta$  is the Laplace operator, the relations are obtained,

$$\frac{\partial}{\partial x_2} \Delta \phi = 0 \quad \text{and} \quad \frac{\partial}{\partial x_1} \Delta \phi = 0 \quad (64 \cdot 5)$$

which require that

$$\Delta \phi = \left( \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right) \phi = c = \text{const} \quad (64 \cdot 6)$$

The boundary condition along the stress-free perimeter,

$$t_1 l_1 + t_2 l_2 = 0 \quad (64 \cdot 7)$$

where  $l_1$  and  $l_2$  are the directional cosines of the boundary element  $ds$  is transformed into the equation,

$$t_1 \frac{dx_2}{ds} - t_2 \frac{dx_1}{ds} = \frac{\partial \phi}{\partial x_2} \cdot \frac{dx_2}{ds} + \frac{\partial \phi}{\partial x_1} \cdot \frac{dx_1}{ds} = 0 \quad (64 \cdot 8)$$

by introducing eqs. 64·3 into 64·7. Thus, along the perimeter  $\phi = \text{const}$ ; this constant may be assumed as zero without changing the stresses. Hence, the problem has been reduced to that of the determination of a potential function for the cross section that is zero along the perimeter. From the consideration of compatibility of strains the constant  $c$  is obtained  $c = 2G\theta$ , where  $\theta$  denotes the angle of twist per unit length. The torque,

$$M = - \iint (t_1 x_2 - t_2 x_1) dx_1 dx_2 = 2 \iint \phi dx_1 dx_2 = 2V \quad (64 \cdot 9)$$

is obtained by introducing eq. 64·3 and by partial integration, considering the boundary condition  $\phi = 0$ .

Comparing eq. 64·6 with that of a thin elastic membrane under pressure, Prandtl<sup>64·2</sup> has suggested that  $\phi$  can be represented as the surface of a membrane under internal pressure; the contour lines are the trajectories of shear stress, and the stress values are proportional to the slope of the membrane. The torque is twice the volume  $V$  of the stress surface  $\phi(x_1 x_2)$ .

In the plastic problem, eqs. 64·2 and 64·3, together with the Huber–Mises–Hencky yield condition,

$$t^2 = t_1^2 + t_2^2 = k^2 \quad (64 \cdot 10)$$

where  $k$  denotes the yield stress in shear, result in the equation:<sup>64·3</sup>

$$\left(\frac{\partial \phi}{\partial x_1}\right)^2 + \left(\frac{\partial \phi}{\partial x_2}\right)^2 = k^2 = |\text{grad } \phi|^2 \quad (64 \cdot 11)$$

This is the equation of a surface with constant slope and can, as suggested by Nadai,<sup>64·3</sup> be represented by a surface of constant slope such as sand heap. By comparing this surface with that of Prandtl's elastic membrane a clear picture of the transformation from the elastic into the plastic state is obtained. By increasing the pressure on the elastic membrane a condition of

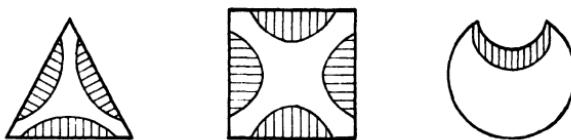


FIG. 64·1 Isotropically plastic areas in torsion.

insetting isotropic yield is reached when the first contact is established between the membrane and the surface of constant slope. Under increased pressure the area of free deformation of the membrane (elastic domain) is gradually reduced. This is shown for different shapes of the cross section in Fig. 64·1. The shaded areas are those in which homogeneous isotropic plastic conditions have been reached. However, since in certain metals such as mild steel the real initiation of yield is heterogeneous, glide lines can, in these metals, be observed to precede any extension of the homogeneous plastic area, as indicated in Fig. 40·1.

The relation 64·9 between torque and volume of the membrane  $\phi$ , derived for the elastic problem, also holds for the plastic problem of torsion. Thus, for a circular shape of radius  $a$  the volume  $V_{pl}$  of the plastic function

$$V_{pl} = \pi a^2 \frac{h}{3} \quad \text{where} \quad \frac{\partial \phi}{\partial r} = k \quad (64 \cdot 12)$$

Hence, the height  $h = ka$  and the torque in the fully plastic state of the section,

$$M_{pl} = 2V_{pl} = 2\pi ka \frac{a^2}{3} = \frac{2}{3}k\pi a^3 \quad (64 \cdot 13)$$

The elastic torque is given by

$$M = \frac{\pi a^3 t}{2} \quad (64 \cdot 14)$$

which, for  $t = k$ , attains the limiting elastic value,

$$M_{EI} = \frac{1}{2} k \pi a^3 = \frac{3}{4} M_{Fl} \quad (64 \cdot 15)$$

For the equilateral triangle section,

$$M_{Fl} = \frac{1}{12} k a^3 \quad \text{and} \quad M_{EI} = \frac{1}{20} k a^3 = \frac{3}{5} M_{Fl} \quad (64 \cdot 16)$$

and, for the square section,

$$M_{Fl} = \frac{1}{3} k a^3 \quad \text{and} \quad M_{EI} = \frac{1}{4.81} k a^3 = 0.625 M_{Fl} \quad (64 \cdot 17)$$

If a torque  $M_{EI} < M < M_{Fl}$  is applied and released, a state of residual stresses is introduced; these stresses are obtained by subtracting, from the elastic-plastic stresses produced by the torque  $M$ , the elastic stresses that would be produced by the same value of the torque.

## 65. Bending. Nonuniform Stress

A problem of elastic-plastic equilibrium that is of considerably practical interest and that by the introduction of the Bernoulli-Navier assumption can be simplified and formulated as a prob-

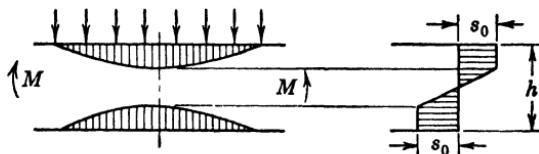


FIG. 65·1 Isotropically plastic areas in pure bending and stress distribution over cross section.

lem of uniaxial nonhomogeneous stress is that of elastic-plastic bending. When a member is loaded in bending, a plastic zone will develop near the edges of the cross section carrying the maximum bending moment and will gradually extend into the interior under increasing load until it finally reaches the neutral axis (Fig. 65·1). This state delimits the conditions of contained

plastic deformation and of free plastic flow. Again, the assumption of homogeneous extension of isotropic plastic areas is frequently at variance with the real process of initiation of plastic deformation and its progress, at least in the early stages. The stage during which isotropic plastic domains are formed is preceded by a stage of heterogeneous plastic deformation characterized by the appearance of glide lines as shown in Fig. 40·1.

Under the assumption of isotropic plastic deformation under constant stress the linear stress distribution over the cross section in problems of elastic bending is replaced by the broken distribution shown in Fig. 65·1. Extending the Bernoulli–Navier assumption of a plane section remaining plane after deformation to elastic–plastic deformations (Fig. 65·2), the statical

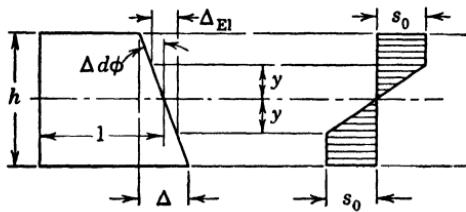


FIG. 65·2 Deformation and stress distribution of elastic–plastic rectangular section in pure bending.

moment  $M_i$  of the stresses with respect to the neutral axis is given by the equation,<sup>65·1</sup>

$$M_i = EI\Delta d\phi = EI \frac{\Delta}{h} \quad (65\cdot1)$$

where  $I$  denotes the moment of inertia of the cross section, if this section is symmetrical with respect to the neutral axis. The statical moment of the stresses associated with yield initiation is given by

$$M_{EI} = s_0 \frac{2I}{h} \quad (65\cdot2)$$

where  $s_0$  denotes the yield stress in uniaxial tension or compression. The absolute sum of extreme fiber extension and compression  $\Delta_{EI}$  pertaining to  $M_{EI}$  is given by

$$\Delta_{EI} = 2 \frac{s_0}{E} = 2 \frac{s_0}{3G} \quad (65\cdot3)$$

If  $M$  denotes the applied external bending moment, the equilibrium condition is expressed by  $M = M_i$ . Purely elastic parts are defined by  $M < M_{EI}$ , whereas values of  $M > M_{EI}$  characterize domains of elastic-plastic deformation. Within those domains, symmetric with regard to the neutral axis, regions of plastic deformation extend from the outer fibers toward the interior; the height of the plastic regions is  $(h/2 - y_0)$ ,  $y_0$  being one-half the height of the remaining elastic core. If the deformed section remains plane,

$$2y = \frac{\Delta_{EI}}{\Delta} h \quad (65 \cdot 4)$$

The statical moment of the stresses in the elastic-plastic state

$$M_i = \int_{-h/2}^{+h/2} s y dA = \int_0^{y_0} s y dA + 2s_0 \int_{y_0}^{h/2} y dA \quad (65 \cdot 5)$$

For a rectangular section, considering eq. 65·4 and 65·1,

$$\begin{aligned} M_i &= \frac{3}{2} M_{EI} \left[ 1 - \frac{1}{3} \left( \frac{\Delta_{EI}}{\Delta} \right)^2 \right] \\ &= \frac{3 EI}{2 h} \Delta_{EI} \left[ 1 - \frac{1}{3} \left( \frac{\Delta_{EI}}{\Delta} \right)^2 \right] = M \end{aligned} \quad (65 \cdot 6)$$

Hence

$$\Delta = \pm \Delta_{EI} \sqrt{1 - \frac{2M}{3M_{EI}}} \quad (65 \cdot 7)$$

and, for a known distribution  $M(x)$ , the equation of the elastic-plastic boundary,

$$y_0 = \pm \frac{h}{2} \sqrt{3 - 2 \frac{M(x)}{M_{EI}}} \quad (65 \cdot 8)$$

Introducing the origin of the coordinate system at mid-span, the moment equation of a freely supported beam of span  $l$  with uniform load  $p$ ,

$$M = \frac{1}{8} pl^2 \left[ 1 - 4 \left( \frac{x}{l} \right)^2 \right] \quad (65 \cdot 9)$$

By introducing eq. 65·9 into eq. 65·8 the boundary of the plastic area is obtained after a short transformation,

$$y_0 = \pm \frac{h}{2} \left[ \left( 3 - 2 \frac{p}{p_{EI}} \right) + 8 \frac{p}{p_{EI}} \left( \frac{x}{l} \right)^2 \right]^{\frac{1}{2}} \quad (65 \cdot 10)$$

where  $p_{EI}$  denotes the load pertaining to  $M_{EI}$ . Introducing

$$\frac{2y}{h} = \eta, \quad \frac{x}{l} = \xi, \quad \left( 3 - 2 \frac{p}{p_{EI}} \right) = A$$

and

$$\left( 3 - 2 \frac{p}{p_{EI}} \right) \frac{p_{EI}}{8p} = B \quad (65 \cdot 11)$$

eq. 65·10 can be written in the form,

$$\frac{\eta^2}{A} - \frac{\xi^2}{B} = 1 \quad (65 \cdot 12)$$

which is the equation of a hyperbola, the asymptotes of which are defined by the condition  $p_{EI} = \frac{M_{i_{\max}}}{M_{EI}} p_{EI} = 1.5 p_{EI}$ , since, according to eq. 65·6,  $\frac{3}{2} M_{EI}$  is the maximum limiting moment a rectangular section can sustain under conditions of contained plastic deformation. For  $p = 1.5 p_{EI}$  the equation of the hyperbola degenerates into the equation of the asymptotes:

$$\eta^2 - 12\xi^2 = 0 \quad \text{or} \quad \eta = \pm \xi \sqrt{12} \quad (65 \cdot 13)$$

The plastic area of a freely supported beam under uniformly distributed load is shown in Fig. 65·3.

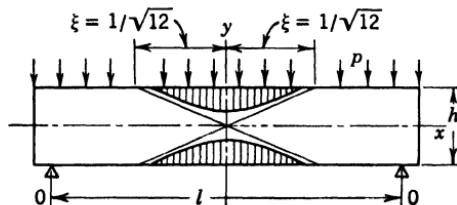


FIG. 65·3 Plastic areas of rectangular beam subject to uniformly distributed load.

For a simply supported beam with a concentrated load at mid-span,

$$M = \frac{1}{4}Pl \left( 1 \pm \frac{2x}{l} \right) \quad (65 \cdot 14)$$

and the equation of the elastic-plastic boundary,

$$\eta^2 = \pm 4\xi + \left( 3 - 2 \frac{P}{P_{EI}} \right) \quad (65 \cdot 15)$$

This is the equation of two parabolas approaching each other as the load increases. Under the limiting load  $P_{fl} = 1.5P_{EI}$  their vertices touch at mid-span (Fig. 65·4).

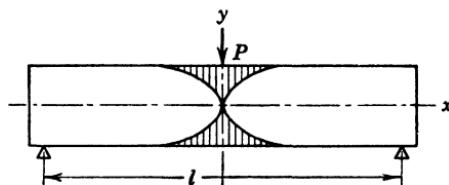


FIG. 65·4 Plastic areas of rectangular beam under maximum concentrated load.

If the cross section of the bent bar is not symmetrical with respect to the neutral axis, the procedure of computation of the elastic-plastic boundary is similar; it is only the computation of  $M_i$  by integration of eq. 65·5 that becomes more elaborate.<sup>65·2</sup>

On release of the load that has produced a moment  $M_{EI} < M < M_{fl}$ , a state of residual stresses remains.

The distribution of residual stresses is obtained by subtracting from the elastic-plastic stresses due to  $M$  a state of elastic stresses pertaining to the same moment  $M$  (Fig. 65·5). Because of the equilibrium condition  $\int sy \, dA = 0$ , residual stresses in bent sections are both tensile and compressive; the boundary between the regions of tensile and compressive residual stresses are the points of intersection of the elastic-plastic load-

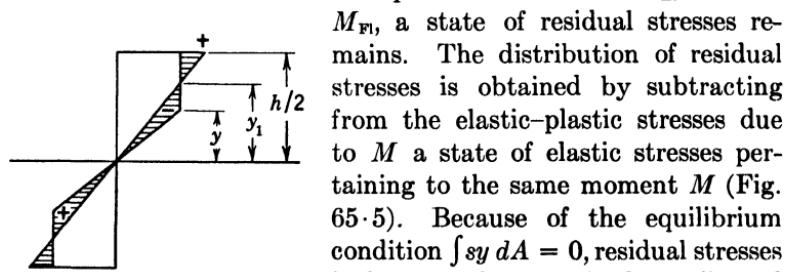


FIG. 65·5 Residual bending stresses in symmetrical section previously strained beyond the yield limit.

ing and the elastic unloading stress distribution, obtained from the relation,

$$s_{\text{unloading}} = \frac{M}{I} y_1 = s_0 \quad (65 \cdot 16)$$

Hence, for a rectangular section, considering eq. 65·6,

$$\eta_1 = \frac{2}{h} s_0 \frac{I}{M} = \frac{4}{3} s_0 \frac{I}{h M_{EI}} \frac{1}{1 - \frac{1}{3}\eta^2} \quad (65\cdot17)$$

This is the equation of the ordinates  $\eta_1 = 2y_1/h$  of zero residual stress in terms of that of the preceding elastic-plastic boundary.

Introducing eq. 65·4 into eq. 65·6 the ratio between the elastic-plastic resisting moment  $M_i$  and the limiting elastic moment  $M_{EI}$  is obtained as a function of the depth of penetration of the plastic zone  $c = (h/2) - y$ :

$$M_i/M_{EI} = \frac{3}{2} \left[ 1 - \frac{1}{3} \left( 1 - \frac{2c}{h} \right)^2 \right] \quad (65\cdot18)$$

This function, which is represented in Fig. 65·6, is a parabola with the vertex  $M_i/M_{EI} = 1.5$  at  $c = h/2$ ; its tangent at the point  $M_i/M_{EI} = 1.0$  and  $c = 0$  has an inclination of

$$\begin{aligned} \tan \alpha &= \frac{d}{dc} (M_i/M_{EI}) \\ &= \frac{2}{h} \left( 1 - \frac{2c}{h} \right)_{c=0} = \frac{2}{h} \end{aligned} \quad (65\cdot19)$$

If we denote by  $K$  the ratio of the maximum fiber stress to the mean stress on either side of the neutral axis and by  $S$  the inclination of the stress distribution, the limiting elastic stress distribution in simple bending is defined by  $K = 2$  and  $S = 2s_0/h$ , the mean stress being  $s = s_0/2$ . The inclination of the tangent to the function  $M_i/M_{EI} = f(c/h)$  at  $c = 0$ , as given by eq. 65·19, which indicates the rate of increase of the resisting moment within the region of contained small plastic deformations, can thus be expressed in the form:

$$\tan \alpha = \frac{2}{h} = \frac{S}{s_{max}} = \frac{S}{K\bar{s}} = \frac{S}{s_0} \quad (65\cdot20)$$

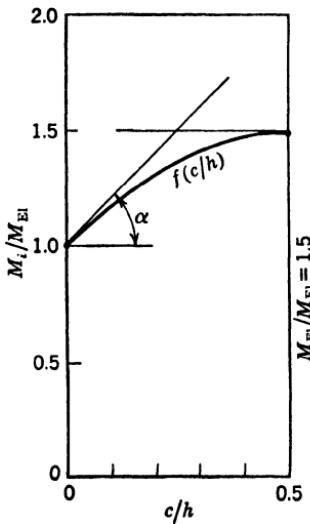


FIG. 65·6 Relation between elastic-plastic resisting moment  $M_i$  of rectangular section in bending and depth of penetration  $c$  of plastic zone.



to a bending moment  $M$  combined with an axial force  $P$  is obtained from the limiting stress distribution (Fig. 65·7) by evaluating the equilibrium conditions:

$$\int s dA = P \quad \text{and} \quad \int sy dA = M \quad (65 \cdot 21)$$

For a rectangular section this evaluation leads to the relations:

$$s_0(2d - h) = P \quad \text{and} \quad s_0dc = M \quad (65 \cdot 22)$$

For an axial force alone  $d = h$ ; for pure bending  $c = d = h/2$ . Hence,

$$P_{\text{Pl}} = s_0h = P_{\text{El}}$$

$$\text{and} \quad M_{\text{Pl}} = \frac{s_0h^2}{4} = 1.5M_{\text{El}} \quad (65 \cdot 23)$$

Dividing the eqs. 65·22 by the respective eqs. 65·23 yields

$$(P/P_{\text{Pl}}) = 2d/h - 1$$

$$\text{and} \quad (M/M_{\text{Pl}}) = 4cd/h^2 \quad (65 \cdot 24)$$

By combining these two equations the critical relation is obtained between  $(P/P_{\text{Pl}})$  and  $(M/M_{\text{Pl}})$  which defines the combinations  $(M/P)$  of axial force and bending moment under which the limiting plastic carrying capacity of a rectangular section is attained:

$$(M/M_{\text{Pl}}) = 1 - (P/P_{\text{Pl}})^2 \quad (65 \cdot 25)$$

The limiting elastic carrying capacity is defined by the relation

$$(P/P_{\text{El}}) + (M/M_{\text{El}}) = (P/P_{\text{Pl}}) + \frac{3}{2}(M/M_{\text{Pl}}) = \pm 1 \quad (65 \cdot 26)$$

The relations 65·25 and 65·26 are represented in Fig. 65·8. The area between the two functions represents states of elastically contained plastic deformation under combined bending and axial loading.

The validity of all foregoing equations for the elastically contained plastic bending, based on an isotropic condition of plasticity, has been repeatedly questioned on the basis of results of

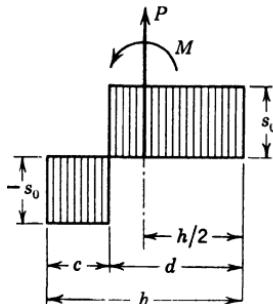


FIG. 65·7 Limiting plastic stress distribution under combined bending moment and axial force.

bending tests of various steel shapes, which suggest that the transformation from elastic to plastic bending does not take place by the spreading of a homogeneously plasticized region under a constant stress  $s_0$ , but rather by the sudden breakdown of an elastic state in which the extreme fiber stress has been able to attain a value  $s_{01} > s_0$ . The stress  $s_{01}$  is considered to represent the upper yield limit of the material the excess of which over  $s_0$  produces the delay in the slip initiation and the

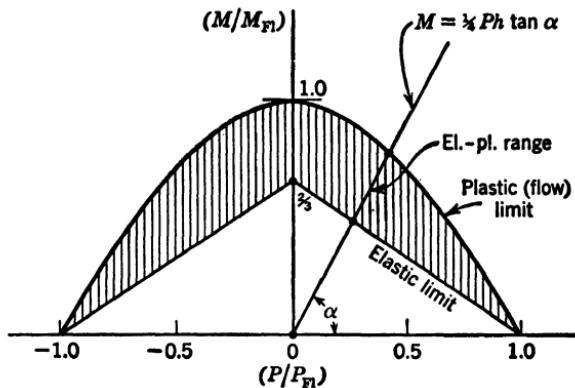


FIG. 65·8 Range of elastically contained plastic deformation of rectangular section under combined bending moment and axial force.

consequent raising of  $M_{EI}$  above the value based on yield initiation at a stress  $s_0$ . Because of this delay plastic deformation is no longer an isotropic process but consists of a series of sudden glide processes involving successive layers, which start after the stresses over a considerable depth of the section have exceeded  $s_0$ . The assumption is thus introduced that yield initiation does not depend on the local stress in the extreme fibers alone but on the stress gradient in the vicinity of the considered point, and, thus on the entire stress field.<sup>65·4</sup>

Concerning the critical bending moment under which the process of gliding in layers is initiated, two different assumptions have been made. One of them considers that slip may be delayed until the elastic moment reaches the full plastic value  $M_{Fl}$ , when the resistance of the section breaks down by excessive yielding.<sup>65·5</sup> According to the second assumption, which is in better agreement with test results, the breakdown by sudden heterogeneous yielding occurs under a bending moment which is lower than

$M_{\text{fl}}$  and which is obtained from the purely empirical assumption that for this bending moment the stress ordinate which divides the "stress prism" into two parts of equal volume attains the value  $s_0$ ; this stress ordinate represents the *mean resistance*. The stress prisms are made up of the stress ordinates over the basis formed by the parts of the cross section on either side of the neutral axis (Fig. 65·9). The second assumption is based on the consideration that the "overstressed" region of the stress prism in which  $s > s_0$  is supported by the "understressed" region in which  $s < s_0$  and that, therefore, for this support to

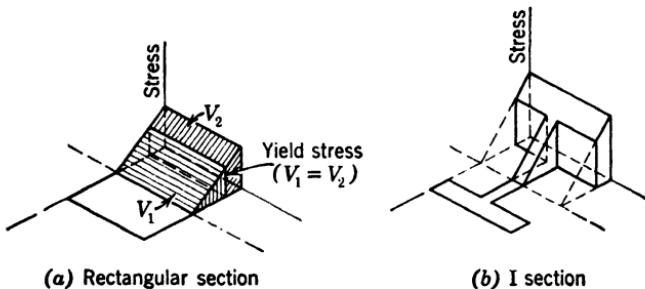


FIG. 65·9 Schematic representation of stress prisms and determination of yield stress in bending for rectangular section.

be effective, the total force in the overstressed region should not exceed the force in the understressed region. Although this consideration has hardly any physical basis, it has been found to represent certain test results fairly well.

The difference between the two assumptions does not affect the numerical results very significantly. For the rectangular cross section the first assumption would give  $s_{01} = 1.5s_0$  since the full plastic moment  $M_{\text{fl}} = 1.5M_{\text{EI}}$ , whereas the second assumption leads to a raised yield stress in bending  $s_{01} = 1.41s_0$ .

There exists considerable evidence that the initiation of plastic deformation, particularly for mild-steel sections, is actually a process of gliding in layers, starting frequently with considerable delay after the application of the load and, for constant load, proceeding at a decreasing rate until a new equilibrium position has established itself.<sup>66-6</sup> However, this evidence does not furnish a conclusive reply to the question of whether the yield stress in bending is raised as a result of the stress gradient or whether the excess of  $s_{01}$  over the yield stress  $s_0$  is a manifesta-

tion of a real upper yield point of the steel accompanied by slip delay, which would therefore be a characteristic property of certain steels, particularly steels of the strain-aging type (see Art. 20).

Whatever the results of future tests concerning the raised yield stress under which plastic deformation in bending is initiated, it appears certain that, because of the heterogeneity of the glide processes associated with plastic deformation, the fully plastic resistance moment  $M_{fl}$ , in general, will not be attained since glide layers will reach the neutral axis in advance of the quasi-isotropic gradually spreading plastic regions. This consideration is of importance in developing design methods based on elastically contained plastic deformation (see Art. 80).

## 66. The Thick-Walled Cylinder under External and Internal Pressure

This problem has been investigated and solved under various simplifying assumptions by a number of authors.<sup>66·1</sup> The procedure outlined here follows essentially that recently developed by the Russian investigators<sup>66·2</sup> on the basis of the Haar-Kármán-Hencky equations of plastic deformation. In solving the problem cylindrical coordinates are introduced. Because of the condition of polar symmetry the general compatibility relations are simplified and only one single equilibrium condition remains.

The problem is geometrically defined by the internal radius of the cylinder  $r_i$ , its external radius  $r_e$ , and the radius delimiting the elastic and the plastic regions  $r_0$ , as expressed by the ratios  $\rho = \frac{r}{r_i}$ ,  $\rho_e = \frac{r_e}{r_i}$ , and  $\rho_0 = \frac{r_0}{r_i}$ . The acting internal pressure is  $q_i$ , the acting external pressure  $q_e$ .

The elastic solution of the problem is given<sup>66·3</sup> by the expressions for the stress components  $s_r$  and  $s_\theta$ :

$$s_r = \frac{\rho_e^2(q_e - q_i) + \rho^2(q_i - q_e\rho_e^2)}{\rho^2(\rho_e^2 - 1)} \quad (66\cdot1)$$

and

$$s_\theta = \frac{-\rho_e^2(q_e - q_i) + \rho^2(q_i - q_e\rho_e^2)}{\rho^2(\rho_e^2 - 1)} \quad (66\cdot2)$$

In the case of the short open cylinder (plane stress)  $s_z = 0$ ; hence,

$$p = \frac{2(q_i - q_e \rho_e^2)}{3(\rho_e^2 - 1)} \quad (66 \cdot 3)$$

In the case of the infinitely long cylinder (plane strain)  $e_z = 0$ ; hence,

$$2Ge_z = s_z - \frac{3\mu}{1 + \mu} p = 0 \quad (66 \cdot 4)$$

and

$$s_z \mu (s_r + s_\theta) = \mu \frac{2(q_i - q_e \rho_e^2)}{(\rho_e^2 - 1)} = \frac{3\mu}{1 + \mu} p \quad (66 \cdot 5)$$

Thus, for the infinitely long cylinder,

$$p = \frac{2(1 + \mu)(q_i - q_e \rho_e^2)}{3(\rho_e^2 - 1)} \quad (66 \cdot 6)$$

The longitudinal stress  $s_z$  in the infinitely long cylinder under internal pressure,

$$s_z = 2\mu + \frac{q_i}{\rho_e^2 - 1} \quad (66 \cdot 7)$$

In the closed cylinder under internal pressure,

$$s_z = \frac{q_i r_i^2 \pi}{\pi(r_e^2 - r_i^2)} = \frac{q_i}{\rho e^2 - 1} \quad (66 \cdot 8)$$

Comparison of eqs. 66·7 and 66·8 shows that for volume-constant deformation the states of stresses and strain in both the infinitely long and the closed cylinder under internal pressure are identical.

The limiting values of either the internal or the external pressure which initiates a plastic ring along the inner surface of the cylinder is obtained by introducing the stress components into the Huber-Mises-Hencky yield condition. Hence, for the short cylinder,

$$q_{eEI} = s_0 \frac{\rho_e^2 - 1}{\sqrt{3\rho_e^4 + 1}}; \quad q_e = 0 \quad (66 \cdot 9)$$

and

$$q_{eEI} = s_0 \frac{\rho_e^2 - 1}{2\rho_e^2}; \quad q_i = 0 \quad (66 \cdot 10)$$

For the infinitely long cylinder,

$$q_{iEI} = s_0 \frac{(\rho_e^2 - 1)}{\sqrt{(1 - 2\mu)^2 + 3\rho_e^4}}; \quad q_e = 0 \quad (66 \cdot 11)$$

and

$$q_{eEI} = s_0 \frac{(\rho_e^2 - 1)}{2\rho_e^2 \sqrt{1 - \mu + \mu^2}}; \quad q_i = 0 \quad (66 \cdot 12)$$

The solution for the fully plastic state of a cylinder of external radius  $r_e = r_0$ , which is attained when the boundary of the elastic zone has been pushed back to  $\rho = \rho_0 = \rho_e$  is relatively simple only if the plastic material is assumed to be incompressible ( $\mu = 0.5$ ). Under this assumption the strain components satisfy the equation  $e_v = 0$ ; considering the compatibility eqs. 26·5, this equation may be written in the form,

$$\frac{du_r}{dr} + \frac{u}{r} + a = 0 \quad (66 \cdot 13)$$

which, integrated, gives the equation for  $u_r$ ,

$$u_r = \frac{C}{r} - \frac{ar}{2} \quad (66 \cdot 14)$$

where  $C$  is an integration constant. Defining this constant arbitrarily by

$$C = \frac{cs_0r_0^2}{2G} \quad (66 \cdot 15)$$

where  $c$  is another constant, the radial displacement,

$$u_r = \frac{cs_0\rho_0^2 - aG\rho^2}{2G\rho} r_i \quad (66 \cdot 16)$$

Therefore, the strains,

$$e_r = -\frac{cs_0\rho_0^2 + aG\rho^2}{2G\rho^2} \quad (66 \cdot 17)$$

$$e_\theta = \frac{cs_0\rho_0^2 - aG\rho^2}{2G\rho^2}$$

By introducing eqs. 66·17 into the Hencky eq. 42·15, the stress-strain relations are obtained:

$$\begin{aligned}s_\theta - s_r &= \frac{2G}{1 + \phi} (e_\theta - e_r) = \frac{2cs_0\rho_0^2}{(1 + \phi)\rho^2} \\s_\theta - s_z &= \frac{2G}{1 + \phi} (e_\theta - e_z) = \frac{cs_0\rho_0^2 - 3Gap^2}{(1 + \phi)\rho^2} \quad (66\cdot18) \\s_r - s_z &= \frac{2G}{1 + \phi} (e_r - e_z) = \frac{-cs_0\rho_0^2 - 3Gap^2}{(1 + \phi)\rho^2}\end{aligned}$$

By substituting these relations into the yield condition, an expression for the function  $\phi$  is obtained:

$$(1 + \phi) = \frac{\sqrt{3}}{s_0\rho} \sqrt{c^2s_0^2\rho_0^4 + 3G^2a^2\rho^4} \quad (66\cdot19)$$

Introducing the first of the eq. 66·18 into the equilibrium condition 26·2 and considering that  $r/dr = \rho/d\rho$ , the differential equation for  $s_r$  is obtained

$$\rho \frac{ds_r}{d\rho} = \frac{2cs_0\rho_0^2}{(1 + \phi)\rho^2} = \frac{2cs_0^2\rho_0^2}{\sqrt{3} \sqrt{c^2s_0^2\rho_0^4 + 3G^2a^2\rho^4}} \quad (66\cdot20)$$

the integration of which gives

$$s_r = \frac{s_0}{2\sqrt{3}} \log \left[ \frac{(c^2s_0^2\rho_0^4 + 3G^2a^2\rho^4)^{\frac{1}{2}} - s_0c\rho_0^2}{(c^2s_0^2\rho_0^4 + 3G^2a^2\rho^4)^{\frac{1}{2}} + s_0c\rho_0^2} \right] + D \quad (66\cdot21)$$

where  $D$  is an integration constant.

The constants  $a$ ,  $D$ , and  $c$  are defined by the boundary conditions:

(a)  $s_r = -q_i$  for  $r = r_i$ , and  $s_r = 0$  for  $r = r_e$  for the cylinder under internal pressure, or  $s_r = 0$  for  $r = r_i$ , and  $s_r = -q_e$  for  $r = r_e$  for the cylinder under external pressure, and

(b) The condition that the integral over the longitudinal stresses  $s_z$  equals the externally applied longitudinal force which is zero for the open cylinder and  $P = q_e 2\pi r_i^2$  for the closed cylinder.

If the elastic-plastic state is considered and the boundary between the elastic and the plastic region defined by  $r = r_0$ ,

the following conditions must be imposed along this boundary to fit the solution for the elastic and the plastic state continuously:

1. Stress components  $s_r, s_\theta, s_z$  continuous.
2. Strain components  $e_r, e_\theta, e_z$  continuous.
3.  $\phi = 0$ .

By introducing  $\rho = \rho_0$  and  $(1 + \phi) = 1$  into eq. 66·19 the constant  $a$  is obtained:

$$a = e_z = -\frac{s_0}{3G} \sqrt{1 - 3c^2} \quad (66 \cdot 22)$$

As a result of the equality of the radial stresses  $s_r = s_{r0}$ , where  $s_{r0}$  denotes the stress along  $r = r_0$ , the integration constant  $D$  follows from eq. 66·21 with  $\rho = \rho_0$ :

$$D = s_{r0} - \frac{s_0}{2\sqrt{3}} \log \frac{1 - c\sqrt{3}}{1 + c\sqrt{3}} \quad (66 \cdot 23)$$

The plastic solution is considerably simplified for the infinitely long or the closed volume-constant cylinder ( $e_z = 0$ ). Introducing this condition, eq. 66·22 can be used directly to determine the constant  $c = 1/\sqrt{3}$ . With  $a = 0$  the equilibrium equation becomes

$$\rho \frac{ds_r}{d\rho} = \frac{2s_0}{\sqrt{3}} \quad (66 \cdot 24)$$

which, integrated, gives the simplified eq. 66·21 for the radial stress,

$$s_r = \frac{2s_0}{\sqrt{3}} \log \rho + D \quad (66 \cdot 25)$$

Because of the boundary conditions  $s_r = s_{r0}$  for  $\rho = \rho_0$ ,

$$D = s_{r0} - \frac{2s_0}{\sqrt{3}} \log \rho_0 \quad (66 \cdot 26)$$

Therefore,

$$s_r = s_{r0} + \frac{2s_0}{\sqrt{3}} \log \left( \frac{\rho}{\rho_0} \right) \quad (66 \cdot 27)$$

With  $s_r = -q_i$  along  $r = r_i$  ( $q_e = 0$ ) eq. 66·27 becomes

$$q_i = -s_{r0} - \frac{2s_0}{\sqrt{3}} \log \left( \frac{\rho}{\rho_0} \right) \quad (66 \cdot 28)$$

A similar expression is obtained when  $s_r = -q_e$  for  $r = r_e$  ( $q_i = 0$ ). Since

$$s_\theta = s_{r0} + \frac{2s_0}{\sqrt{3}} \log \left( \frac{\rho}{\rho_0} \right) + 1 \quad (66 \cdot 29)$$

for  $\rho = \rho_0$ ,

$$s_{\theta0} = s_{r0} + \frac{2s_0}{\sqrt{3}} \quad (66 \cdot 30)$$

Because of the boundary conditions at  $\rho = \rho_0$  and the condition  $s_r = -q_e = 0$  for  $\rho = \rho_e$ ,

$$s_{\theta0} = s_{r0} \frac{\rho_e^2 + \rho_0^2}{\rho_e^2 - \rho_0^2} \quad (66 \cdot 31)$$

Hence,

$$s_{r0} = - \frac{\rho_e^2 - \rho_0^2}{\rho_e^2} \frac{s_0}{\sqrt{3}} \quad (66 \cdot 32)$$

Thus according to eqs. 66·27 and 66·29, the stresses within the plastic region of the infinitely long cylinder with internal pressure are

$$\begin{aligned} s_r &= \frac{s_0}{\rho_e^2 \sqrt{3}} \left[ 2\rho_e^2 \log \left( \frac{\rho}{\rho_0} \right) - \rho_e^2 + \rho_0^2 \right] \\ s_\theta &= \frac{s_0}{\rho_e^2 \sqrt{3}} \left[ 2\rho_e^2 \log \left( \frac{\rho}{\rho_0} \right) + \rho_e^2 + \rho_0^2 \right] \end{aligned} \quad (66 \cdot 33)$$

and, because of constant volume,

$$s_r = \frac{1}{2}(s_r + s_\theta) = \frac{s_0}{\rho_e^2 \sqrt{3}} \left[ 2\rho_e^2 \log \left( \frac{\rho}{\rho_0} \right) + \rho_0^2 \right] \quad (66 \cdot 34)$$

Within the elastic region,

$$s_{re} = - \frac{s_0 \rho_0^2}{\rho_e^2 \rho^2 \sqrt{3}} (\rho_e - \rho^2) \quad (66 \cdot 35)$$

$$s_{\theta e} = \frac{s_0 \rho_0^2}{\rho_e^2 \rho^2 \sqrt{3}} (\rho_e^2 + \rho^2)$$

and

$$s_{ze} = \mu(s_r + s_\theta) = \mu \frac{2s_0\rho_0^2}{\rho_e^2 \sqrt{3}} \quad (66 \cdot 36)$$

The relation between the internal pressure and the radius ratio  $\rho_0 = r_0/r_i$  of the boundary between the elastic and the plastic zones is obtained from the first of eqs. 66·33 introducing the condition  $s_r = -q_i$  for  $r = r_i$ :

$$q_i = \frac{s_0}{\rho_e^2 \sqrt{3}} (2\rho_e^2 \log \rho_0 + \rho_e^2 - \rho_0^2) \quad (66 \cdot 37)$$

The hydrostatic stress within the plastic region,

$$p = \frac{s_0}{\rho_e^2 \sqrt{3}} \left[ 2\rho_e^2 \log \left( \frac{\rho}{\rho_0} \right) + \rho_0^2 \right] \quad (66 \cdot 38)$$

and, within the elastic region,

$$p_e = \frac{2(1 + \mu)s_0\rho_0^2}{3\rho_e^2 \sqrt{3}} = \frac{s_0}{\rho_e^2 \sqrt{3}} \frac{2(1 + \mu)}{3} \rho_0^2 \quad (66 \cdot 39)$$

For  $e_z = a = 0$ , eq. 66·19 is transformed into

$$(1 + \phi) = \left( \frac{\rho_0}{\rho} \right)^2 \quad \text{and} \quad \phi = \frac{\rho_0^2 - \rho^2}{\rho^2} \quad (66 \cdot 40)$$

Hence the strain components in the plastic region, according to eqs. 66·18,

$$e_r = \frac{s_0\rho_0^2}{2G\rho^2 \sqrt{3}}; \quad e_\theta = \frac{s_0\rho_0^2}{2G\rho^2 \sqrt{3}}; \quad e_z = 0 \quad \text{and} \quad e_v = 0 \quad (66 \cdot 41)$$

The strain components within the elastic region

$$\begin{aligned} e_{re} &= \frac{1}{2G} (s_{re} - p_e) = \frac{s_0\rho_0^2}{2G\rho_e^2 \rho^2 \sqrt{3}} [\rho_e^2 - (1 - 2\mu)\rho^2] \\ e_{\theta e} &= \frac{1}{2G} (s_{\theta e} - p_e) = - \frac{s_0\rho_0^2}{2G\rho_e^2 \rho^2 \sqrt{3}} [\rho_e^2 + (1 - 2\mu)\rho^2] \\ e_z &= 0; \quad e_{ve} = \frac{s_0(1 - 2\mu)\rho_0^2}{G\rho_e^2 \sqrt{3}} \end{aligned} \quad (66 \cdot 42)$$

For the problem of the infinitely long cylinder with external

pressure  $q_e$ , while  $q_i = 0$ , equations similar to those derived previously can be obtained by the same procedure.

Comparison of eqs. 66·41 and 66·42 shows that the boundary conditions for the strains along  $\rho = \rho_0$  are fulfilled only if  $\mu = 0.5$  and  $e_{ve} = e_v = 0$ . For  $\mu \neq 0.5$  the strain components  $e_r$  and  $e_\theta$  in the plastic and the elastic region are incompatible.

The general solution of the problem, not restricted by the assumption of incompressibility of the material, can be obtained

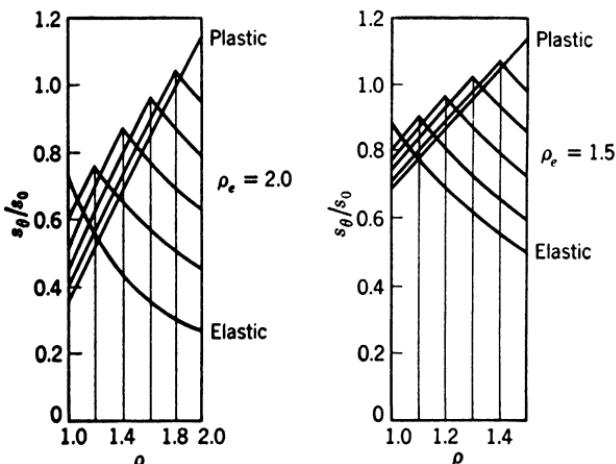


FIG. 66·1 Variation of distribution of tangential stresses in infinitely long thick-walled cylinder under internal pressure (after Beliaev and Sinitski<sup>66·2</sup>).

by the solution for the plastic zone of a system of first-order partial-differential equations resulting from the only equilibrium condition remaining in the problem of rotational symmetry (first of eqs. 26·2), the significant compatibility relations (three left-hand eqs. 26·5), the stress-strain relations (eqs. 42·15 for the deformation theory or eqs. 42·18 for the flow theory), and the condition of plasticity (eq. 41·5 for the Huber-Mises-Hencky condition or eq. 41·21 for the Tresca-St. Venant condition). The boundary conditions of the plastic problem are obtained from the conditions of continuity of stresses and strains along the elastic-plastic boundary. Neither for the flow theory nor for the deformation theory is a solution of the differential equations in closed form possible. The equations must be transformed

cross section of an infinitely long cylinder of wall thickness  $r_i$  and  $0.5r_i$ , respectively, as the diameter of the plastic zone increases under increasing internal pressure.

The problem of elastically contained plastic deformation is transformed into a problem of free plastic flow under a pressure  $q_i$  ( $q_e = 0$ ), or  $q_e$  ( $q_i = 0$ ), for which the elastic zone disappears from the cross section, that is, for  $\rho_0 = \rho_e$ . By introducing this

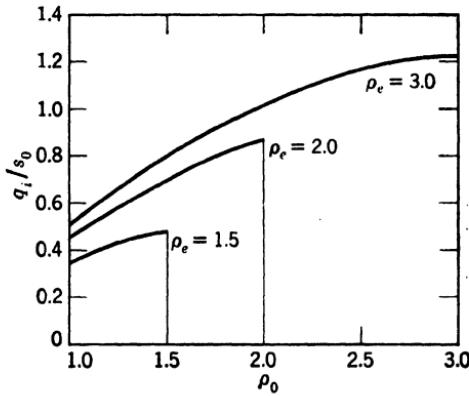


FIG. 66.2 Variation of internal pressure with extension of plastic zone in infinitely long thick-walled cylinder (after Beliaev and Sinitzki<sup>66.2</sup>).

condition into the eqs. 66.33, the stress components in the fully plastic cylinder under internal pressure are obtained:

$$\begin{aligned} s_r &= \frac{2s_0}{\sqrt{3}} \log \left( \frac{\rho}{\rho_e} \right); & s_\theta &= \frac{2s_0}{\sqrt{3}} \left[ \log \left( \frac{\rho}{\rho_e} \right) + 1 \right] \\ s_z &= \frac{s_0}{\sqrt{3}} \left[ 2 \log \left( \frac{\rho}{\rho_e} \right) + 1 \right] \end{aligned} \quad (66.43)$$

The pressure required to produce and maintain plastic yielding over the entire cross section is

$$q_{i\text{fl}} = \frac{2s_0}{\sqrt{3}} \log \rho_e \quad (66.44)$$

Figure 66.2 shows the variation of  $q_i$  with the extension of the plastic zone  $\rho_0$  for cylinders of different wall thickness defined by different values of  $\rho_e$ .

Release of the internal or external pressure  $q_{\text{fl}} < q < q_{\text{m}}$  pro-

duces within the cross section a system of residual stresses. These residual stresses are obtained by subtracting from the stresses of the elastic-plastic condition produced by  $q$  the system of elastic stresses pertaining to the same pressure  $q$ , as defined by eqs. 66·1 and 66·2. Figure 66·3 shows a particular distribution of residual stresses thus obtained in a cylinder of wall thickness  $r_i$ . Because of the residual stresses introduced by prestraining to a pressure  $q_m < q < q_M$ , the cylinder has become elastic for the subsequent application of any pressure up to the prestraining pressure  $q$ . This is true, however, only if within the system of residual stresses itself the yield limit has at no point been exceeded (see Art. 43·2).

The problem of the thick-walled cylinder is transformed into that of a cylindrical hole within an infinite elastic body by introducing  $\rho_e = \infty$ . The stresses within the plastic region of radius  $r_0$  are obtained from the eqs. 66·33 and 66·34.

$$\begin{aligned} s_r &= \frac{s_0}{\sqrt{3}} \left[ 2 \log \left( \frac{\rho}{\rho_0} \right) - 1 \right] \\ s_\theta &= \frac{s_0}{\sqrt{3}} \left[ 2 \log \left( \frac{\rho}{\rho_0} \right) + 1 \right] \quad (66 \cdot 45) \\ s_z &= \frac{2s_0}{\sqrt{3}} \log \left( \frac{\rho}{\rho_0} \right) \end{aligned}$$

Within the elastic region,

$$s_{re} = -s_{\theta e} = -\frac{s_0}{\sqrt{3}} \left( \frac{\rho_0^*}{\rho} \right)^2 \quad (66 \cdot 46)$$

The relation between the internal pressure and the radius of the plastic zone becomes

$$q_i = \frac{s_0}{\sqrt{3}} [2 \log \rho_0 + 1] \quad (66 \cdot 47)$$

These equations first have been derived by Nadai.<sup>66·4</sup>

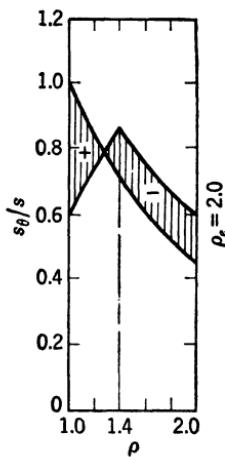


FIG. 66·3 Tangential residual stresses in infinitely long thick-walled cylinder after release of internal pressure  $q_i = 0.68s_0$  (after Beliaev and Sinitzki<sup>66·2</sup>).

## 67. The Rotating Cylinder and Disk

Solutions of the problem of the revolving disk or cylinder have been given by a number of authors, among others by Hencky<sup>67-1</sup> and Donell and Nadai.<sup>67-2</sup> The difference between the equations of the elastic-plastic problem of the thick-walled cylinder and that of the rotating cylinder or disk is in the additional term representing the action of the centrifugal forces in the equilibrium equation. Thus the equilibrium condition,

$$r \frac{ds_r}{dr} - (s_\theta - s_r) = - \frac{w}{\gamma} \omega^2 r^2 \quad (67 \cdot 1)$$

where  $w$  denotes the specific weight of the disk,  $\gamma$  the acceleration of gravity, and  $\omega$  the angular velocity. Under the assumption of constant volume  $e_v = 0$  the radial displacement is obtained by integration of the eq. 66·13 with  $a = e_z$ . Hence, from eq. 66·14 with  $C = 0$  for a solid cylinder and eqs. 26·5,

$$\begin{aligned} u_r &= - \frac{e_z r}{2}; & e_\theta &= \frac{u_r}{r} = - \frac{e_z}{2} \\ e_r &= \frac{du_r}{dr} = - \frac{e_z}{2} = e_\theta \end{aligned} \quad (67 \cdot 2)$$

Since  $e_r - e_\theta = 0$ , the stress difference  $s_r - s_\theta = 0$ . Hence the Huber-Mises-Hencky condition becomes

$$s_\theta - s_z = s_0 \quad (67 \cdot 3)$$

and the equilibrium condition 67·1,

$$\frac{ds_r}{dr} = - \frac{w}{\gamma} \omega^2 r^2 \quad (67 \cdot 4)$$

The radial stress is obtained by integration,

$$s_r = \frac{w}{2\gamma} \omega^2 r_e^2 \left[ 1 - \left( \frac{r}{r_e} \right)^2 \right] = s_\theta \quad (67 \cdot 5)$$

where  $r_e$  denotes the radius of the cylinder. According to eq. 67·3,

$$s_z = s_\theta - s_0 = \frac{w}{2\gamma} \omega^2 r_e^2 \left[ 1 - \left( \frac{r}{r_e} \right)^2 \right] - s_0 \quad (67 \cdot 6)$$

If no force is applied in the direction of the cylinder axis,

$$2\pi \int_0^{r_e} s_z r dr = 0 \quad (67 \cdot 7)$$

or, introducing eq. 67·6,

$$\omega_{\text{Fl}} r_e = 2 \sqrt{\frac{s_0 \gamma}{w}} = v_{e\text{Fl}} \quad (67 \cdot 8)$$

This is the expression for the circumferential speed  $v_e$  at which the cylinder is in a fully plastic state. The pertaining radial and tangential stress components are

$$s_r = s_\theta = 2s_0 \left[ 1 - \left( \frac{r}{r_e} \right)^2 \right] \quad (67 \cdot 9)$$

$$s_z = s_0 \left[ 1 - 2 \left( \frac{r}{r_e} \right)^2 \right]$$

In a perfectly elastic cylinder of radius  $r_e$  revolving with an angular velocity  $v_e = \omega r_e$ , the stress components are expressed by the equations:<sup>67·3</sup>

$$s_{re} = \frac{w v_e^2}{\gamma} \cdot \frac{(3 - 2\mu)}{8(1 - \mu)} \left[ 1 - \left( \frac{r}{r_e} \right)^2 \right] \quad (67 \cdot 10)$$

$$s_{\theta e} = \frac{w v_e^2}{\gamma} \cdot \frac{1}{8(1 - \mu)} \left[ 3 - 2\mu - (1 + 2\mu) \left( \frac{r}{r_e} \right)^2 \right]$$

$$s_{ze} = \frac{w v_e^2}{\gamma} \cdot \frac{\mu}{4(1 - \mu)} \left[ 1 - 2 \left( \frac{r}{r_e} \right)^2 \right]$$

At the center with  $r = 0$ ,  $s_r = s_\theta$ ; hence, according to eq. 67·5, plastic deformation will start when

$$s_\theta - s_z = \frac{(3 - 4\mu)}{8(1 - \mu)} \frac{w v_e^2}{\gamma} = s_0 \quad (67 \cdot 11)$$

or

$$v_e = 2 \sqrt{\frac{2(1 - \mu)}{3 - 4\mu}} \sqrt{\frac{s_0 \gamma}{w}} \quad (67 \cdot 12)$$

Comparison of eqs. 67·8 and 67·12 shows that under the assumption of constant volume ( $\mu = 0.5$ ) the rotating cylinder becomes fully plastic under the peripheral velocity  $v_e = v_{e\text{Fl}}$  without an intermediate elastic-plastic stage. Such a stage is introduced

by values  $\mu \neq 0.5$ ; for the extreme value  $\mu = 0$  the ratio between the peripheral velocity, at which yielding starts at the center and the velocity producing fully plastic conditions reaches its minimum value  $v_e/v_{e\text{fl}} = 0.61$ .

In the case of the rotating disk of constant relatively small thickness, the principal stress  $s_z = 0$ . The two remaining principal stresses should satisfy the yield-condition of the plane-stress problem (eq. 41·19):

$$s_\theta^2 - s_r s_\theta + s_r^2 = s_0^2 \quad (67 \cdot 13)$$

This nonlinear elliptic yield condition leads to solutions of considerable complexity. Since in the problem of the rotating disk the principal stresses  $s_r$  and  $s_\theta$  are both tensile stresses and  $s_\theta > s_r$ , the part of the yield ellipse between  $s_1 = 0$  and  $s_1 = s_0$  in the tensile quadrant (Fig. 41·1) can be replaced in rough approximation by the straight line  $s_2 = s_0$ , which forms one side of the hexagon representing the St. Venant yield condition. Hence the approximate yield condition  $s_\theta = s_0$ .<sup>67·1</sup>

The stresses in the perfectly elastic thin disk of radius  $r_e$ , revolving with an angular velocity  $v_e = \omega r_e$  are expressed by the equations:<sup>67·1</sup>

$$\begin{aligned} s_{re} &= \frac{\omega v_e^2}{\gamma} \cdot \frac{3 + \mu}{8} \left[ 1 - \left( \frac{r}{r_e} \right)^2 \right] \\ s_{\theta e} &= \frac{\omega v_e^2}{\gamma} \cdot \frac{3 + \mu}{8} \left[ 1 - \frac{1 + 3\mu}{3 + \mu} \left( \frac{r}{r_e} \right)^2 \right] \end{aligned} \quad (67 \cdot 14)$$

$$s_z = 0$$

The disk yields first at the center at a velocity obtained from the plasticity condition

$$[s_{\theta e}]_{(r=0)} = \frac{\omega v_e^2}{\gamma} \cdot \frac{3 + \mu}{8} = s \quad (67 \cdot 15)$$

or

$$v_e = \sqrt{\frac{8}{3 + \mu}} \sqrt{\frac{s_0 \gamma}{w}} \quad (67 \cdot 16)$$

By introducing the yield condition  $s_\theta = s_0$  into the equilibrium

eq. 67·1 the differential equation for the radial stress of the fully plastic state is obtained:

$$\frac{d}{dr} (rs_r) = - \frac{w}{\gamma} \omega^2 r^2 + s_0 \quad (67 \cdot 17)$$

Hence,

$$s_r = s_0 - \frac{w}{3\gamma} \omega^2 r^2 \quad \text{and} \quad s_\theta = s_0 \quad (67 \cdot 18)$$

Since for  $r = r_e$  the radial stress  $s_r = 0$ , the circumferential velocity producing fully plastic conditions,

$$v_{e\text{Pl}} = \sqrt{3} \sqrt{\frac{s_0 \gamma}{w}} \quad (67 \cdot 19)$$

Under the assumption of constant volume ( $\mu = 0.5$ ) the ratio  $v_e/v_{e\text{Pl}} = 0.87$ ; under the other extreme assumption  $\mu = 0$ , the ratio  $v_e/v_{e\text{Pl}} = 0.94$ . Hence, the transition between the fully elastic and the fully plastic state of the disk is produced by a relatively small increase of the angular velocity. During this transition the elastic region will extend from a radius  $r = r_0$  to  $r = r_e$ , while a plastic region extends from  $r = 0$  to  $r = r_0$ . The stresses in the elastic region are

$$\begin{aligned} s'_{re} &= \frac{w v_e^2}{\gamma} \cdot \frac{(3 + \mu)}{8} \left( \frac{r}{r_e} \right)^2 + c_1 + \frac{c_2}{r^2} \\ s'_{\theta e} &= \frac{w v_e^2}{\gamma} \cdot \frac{(3\mu + 1)}{8} \left( \frac{r}{r_e} \right)^2 + c_1 + \frac{c_2}{r^2} \end{aligned} \quad (67 \cdot 20)$$

The stresses in the fully plastic region are given by eqs. 67·18. The constants  $c_1$  and  $c_2$  are obtained from the condition  $s'_{re} = s_r$  and  $s'_{\theta e} = s_\theta$  along the elastic-plastic boundary  $r = r_0$ . The relation between the circumferential velocity and the radius of the plastic zone is derived from the condition  $s'_{re} = 0$  for  $r = r_e$ ; this relation is expressed by the equation:

$$\begin{aligned} v_e &= \sqrt{24} \sqrt{\frac{s_0 \gamma}{w}} \left[ 3(3 + \mu) - 2(1 + 3\mu) \left( \frac{r_0}{r_e} \right)^2 \right. \\ &\quad \left. + (1 + 3\mu) \left( \frac{r_0}{r_e} \right)^4 \right] \end{aligned} \quad (67 \cdot 21)$$

The foregoing simplified solution for the elastic-plastic rotating disk is compatible in the stresses but, as in the case of the thick-walled cylinder under pressure, incompatible in strains, unless both the plastic and the elastic regions are considered to be incompressible; however the displacement is continuous across the elastic-plastic boundary.

The introduction of the approximate St. Venant yield condition instead of the correct Huber-Mises-Hencky condition results in differences in the relation between circumferential velocity and the radius of the elastic-plastic boundary which may become considerable near the center of the disk, because of the comparatively high sensitivity of the elastic-plastic state of the disk with respect to small variations in the angular velocity. The difference between the circumferential velocities required to produce a plastic region of a certain radius according to the two conditions increases with increasing radius of the plastic region.

The solution for the rotating disk with a central circular hole of radius  $r_i$  differs from that of the solid disk only by the additional term  $(c_3/r)$  in the expression 67·21 for the radial stress in the plastic zone, which is required in order to comply with the additional boundary conditions  $s_r = 0$  for  $r = r_i$ .

Because of the fact that the axes of principal stress and strain do not rotate in the course of the deformation, problems of rotational symmetry are the only type of problem of elastic-plastic equilibrium for which the flow theories and the deformation theories lead to practically identical solutions if the compressibility of both the elastic and the plastic zone is considered.<sup>67·1</sup>

## 68. The Blunted Wedge

The plastic deformation and resistance of a uniformly loaded blunted wedge was one of the earliest problems investigated by methods of the theory of plasticity. Under conditions of free plastic flow and under certain assumptions concerning the shape of the plastic area (Fig. 68·1), Prandtl has established the following linear relation between the uniform wedge load  $p_w$  necessary to produce and maintain plastic flow and the wedge angle  $\theta$ <sup>68·1</sup>

$$p_w = \text{const}(1 + \theta) \quad (68\cdot1)$$

Results of tests with wedges of various angles  $\theta$  undertaken by Nadai<sup>68-2</sup> and Sachs<sup>68-3</sup> to verify this relation showed discrepancies between theory and tests which increased with the wedge angle  $\theta$ . Only for angles  $\theta$  between 0 and 40° did relation 68-1 describe the test results fairly well; for  $\theta > 40^\circ$  no agreement was found.

The wedge problem illustrates the difference between the two aspects of initiation of plastic deformation: heterogeneous gliding in layers and homogeneous extension of isotropic plastic regions. Prandtl's solution has been derived on the basis of the assumption of unrestricted gliding of the material along glide lines. Nadai's test results showed this assumption to be fairly consistent with real behavior under conditions under which unrestricted gliding is possible. Such conditions exist as long as the glide lines reach the sides of the wedge rather rapidly (Fig. 68-2). In this case the effect of the elastic deformation is negligible, and the gradual extension of isotropic plastic areas is impossible because of the heterogeneous disturbance across the wedge created by the glide lines.

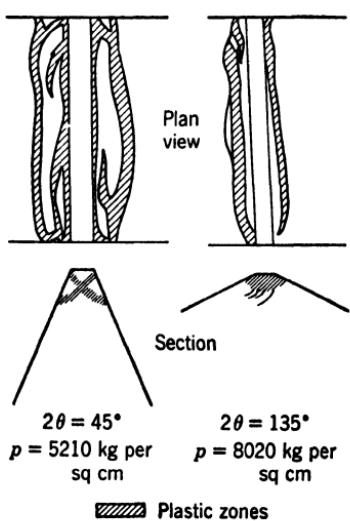


FIG. 68-2 Plastic zones in steel wedges of wedge angle  $2\theta$  under vertical pressure  $p$  (after Nadai<sup>68-2</sup>).

elastic-plastic deformation, the change of which is governed

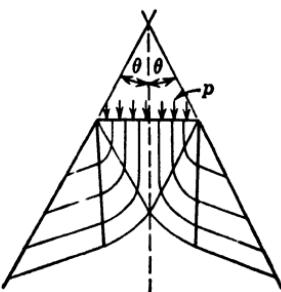


FIG. 68-1 Uniformly loaded blunted wedge under assumption of unrestrained plastic flow (glide lines after Prandtl<sup>68-1</sup>).

conditions under which unrestricted gliding is possible. Such conditions exist as long as the glide lines reach the sides of the wedge rather rapidly (Fig. 68-2). In this case the effect of the elastic deformation is negligible, and the gradual extension of isotropic plastic areas is impossible because of the heterogeneous disturbance across the wedge created by the glide lines.

Such behavior is, however, restricted to comparatively small wedge angles. For large wedge angles the inhomogeneity of the elastic stress field is sufficient to block the progress of the glide lines before they are able to reach the sides of the wedge, and to re-establish a state of contained

essentially by the homogeneous spreading of an isotropic plastic region. Under such conditions Prandtl's assumptions are no longer valid, and this fact is reflected in the increasing discrepancies between relation 68.1 and the test results. According to these results, the load  $p_w$  which produces plastic deformation of different wedges does attain the values predicted by eq. 68.1 for small wedge angles  $\theta$  only; it reaches a limiting maximum value for  $\theta = \pi/2$  which is considerably lower than Prandtl's value  $(p_w)_{\pi/2} = \text{const}(1 + \pi/2)$ .

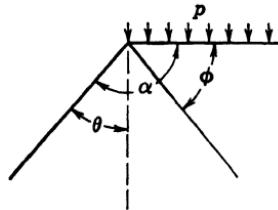


FIG. 68.3 Corner of wedge under uniform load  $p$ .

under uniform load (Fig. 68.3). The stress function of the elastic problem in plane polar coordinates:

$$F = C_1 r^2 + C_2 r^2 \phi + C_3 r^2 \sin 2\phi + C_4 r^2 \cos 2\phi \quad (68.2)$$

The integration constants  $C$  are obtained from the boundary conditions:

$$\begin{aligned} \text{For } \phi = 0, \quad & s_\theta = -p_w, \quad s_{r\theta} = 0 \\ \text{For } \phi = \left(\theta + \frac{\pi}{2}\right), \quad & s_\theta = 0; \quad s_{r\theta} = 0 \end{aligned} \quad (68.3)$$

Introducing

$$A = \frac{1}{2(\alpha - \tan \alpha)}; \quad B = -\frac{1}{2(\alpha \cot \alpha - 1)} \quad (68.4)$$

where  $\alpha = (\theta + \pi/2)$ , the stress components are obtained:

$$\begin{aligned} s_r &= p_w[(B - 1) - 2A\phi - A \sin 2\phi + B \cos 2\phi] \\ s_\theta &= p_w[(B - 1) - 2A\phi + A \sin 2\phi - B \cos 2\phi] \quad (68.5) \\ s_{r\theta} &= p_w[A(1 - \cos 2\phi) - B \sin 2\phi] \end{aligned}$$

Considering a state of plane strain ( $e_z = 0$ ), the Huber-Mises-

Hencky yield condition with the stress components of eq. 68·5 has the form:

$$t_{s\max}^2 = p_w^2[B^2 + 2A^2(1 - \cos 2\phi) - 2AB \sin 2\phi] = \frac{1}{3}s_0^2 \quad (68\cdot6)$$

The maximum value of the principal shear stress  $t_{s\max}$  is reached along a straight line through the corner, defined by the condition  $dt_{s\max}/d\phi = 0$ . With

$$\frac{dt_{s\max}}{d\phi} = p_w^2(4A^2 \sin 2\phi - 4AB \cos 2\phi) = 0 \quad (68\cdot7)$$

or

$$\sin 2\phi - \tan \alpha \cos 2\phi = 0 \quad (68\cdot8)$$

the line of maximum shear  $t_{s\max}$  is defined by  $\phi = \alpha/2$  and thus

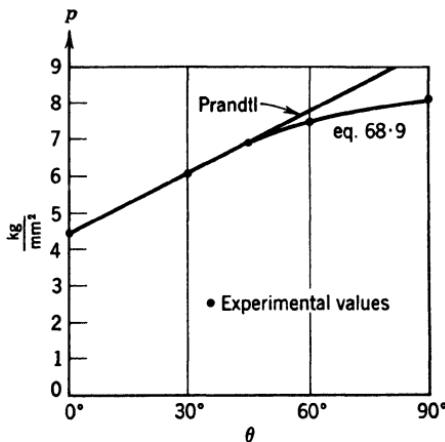


FIG. 68·4 Comparison of yield load of uniformly loaded blunted wedge computed after Prandtl's eq. 68·1 and eq. 68·9 with experimental results.<sup>68·4</sup>

bisects the corner angle. By introducing the value  $\phi = \alpha/2$  into eq. 68·6, the relation between the wedge load  $p_w$  and wedge angle  $\theta$  is obtained:

$$p_w = \frac{2}{\sqrt{3}} s_0 \frac{\left(\theta + \frac{\pi}{2}\right) \sin \theta + \cos \theta}{1 + \sin \theta} \quad (68\cdot9)$$

This relation is presented in Fig. 68·4 together with Prandtl's relation (eq. 68·1) and the test results.

The wedge reaches the fully plastic state when  $t_{s\max} = \frac{1}{\sqrt{3}} s_0$  along  $\phi = 0$  and  $\phi = \alpha$ . Introducing this condition into eq. 68·6, the relation is obtained,

$$t_{s\max}^2 = p_w^2 B^2 = \frac{\tan^2 \alpha}{4(\alpha - \tan \alpha)^2} p_w^2 = \frac{1}{3} s_0^2 \quad (68 \cdot 10)$$

or

$$p_{w\text{fl}} = \frac{2}{\sqrt{3}} s_0 \left[ 1 + \left( \theta + \frac{\pi}{2} \right) \tan \theta \right] \quad (68 \cdot 11)$$

Comparison of eqs. 68·9 and 68·11 shows that the difference in wedge pressure required for yield initiation and for fully plastic flow increases with increasing wedge angle.

## 69. Contact Pressure

With  $\theta = \pi/2$  the problem of the blunted wedge is transformed into the problem of uniform contact pressure (Fig. 69·1) which is of considerable practical importance in the analysis of hardness tests and of the carrying capacity of contact bearings. The assumption of uniform pressure over the contact area is however only one of the possible assumptions and must be modified in accordance with the real conditions.

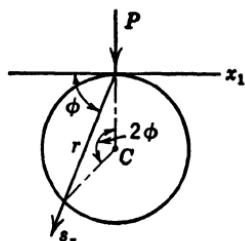
FIG. 69·1 Concentrated force acting on half-plane.

If the contact surface remains plane during deformation, the contact stresses increase towards the edges of the contact area approaching in the elastic problem infinity at the edges. If the contact stresses remain uniform, the elastic deformation of the contact surface increases towards the center.

The elastic solution for any type of contact-pressure distribution can be obtained by integration over a finite contact area of the solution for the single force acting on the half-plane; the stresses for this case expressed in plane-polar coordinates are (Fig. 69·1):

$$s_r = - \frac{2P}{\pi r} \sin \phi; \quad s_\theta = 0; \quad s_{r\theta} = 0 \quad (69 \cdot 1)$$

since the force  $P$  is the resultant of the radial stresses  $s_r$  acting



along any half-circle of radius  $r$ . The maximum shearing stress,

$$t_{\max} = s_r - s_\theta = \text{const} \quad (69 \cdot 2)$$

Hence the lines  $t_{\max} = \text{const}$  are the circles through the point of application of  $P$ , tangent to the  $x_1$  axis. Thus, for a continuously distributed pressure  $p(x_1)$  acting over a finite length of the surface of the half-plane (Fig. 69·2), the

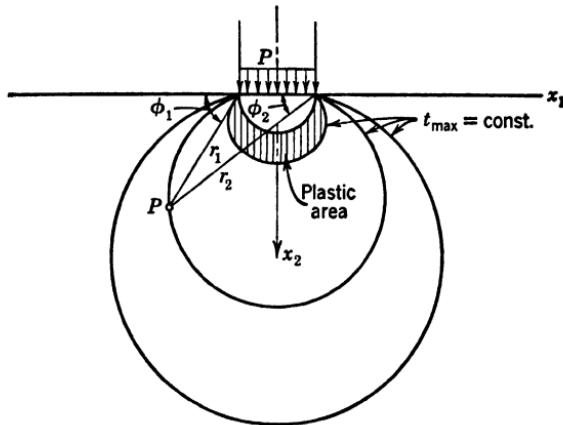


FIG. 69·2 Lines of constant maximum shearing stress  $t_{\max} = \text{const}$  for uniformly distributed contact pressure.

stresses at a point in the interior are determined by integration of eq. 69·1 over this length:

$$\begin{aligned} s_{11} &= \frac{2}{\pi} \int \frac{p(x_1) \cos \phi \sin^2 \phi}{r} dx_1 = - \frac{2}{\pi} \int_{\phi_1}^{\phi_2} p(\phi) \sin^2 \phi d\phi \\ s_{22} &= \frac{2}{\pi} \int \frac{p(x_1) \cos^3 \phi}{r} dx_1 = - \frac{2}{\pi} \int_{\phi_1}^{\phi_2} p(\phi) \cos^2 \phi d\phi \quad (69 \cdot 3) \\ s_{12} &= \frac{2}{\pi} \int \frac{p(x_1) \sin \phi \cos^2 \phi}{r} dx_1 = - \frac{2}{\pi} \int_{\phi_1}^{\phi_2} p(\phi) \sin \phi \cos \phi d\phi \end{aligned}$$

For a uniformly distributed load  $p(\phi) = p_0$ ; hence, by integration,

$$\begin{aligned} s_{11} &= - \frac{p_0}{2\pi} [2(\phi_1 - \phi_2) + \sin 2\phi_1 - \sin 2\phi_2] \\ s_{22} &= - \frac{p_0}{2\pi} [2(\phi_1 - \phi_2) - \sin 2\phi_1 + \sin 2\phi_2] \quad (69 \cdot 4) \\ s_{12} &= \frac{p_0}{2\pi} [\cos 2\phi_1 - \cos 2\phi_2] \end{aligned}$$

The maximum shearing stress,

$$t_{\max} = \left[ \frac{(s_{11} - s_{22})^2}{4} + s_{12}^2 \right]^{\frac{1}{2}} = \frac{p}{\pi} \sin (\phi_1 - \phi_2) \quad (69 \cdot 5)$$

The lines  $t_{\max} = \text{const}$  are therefore circles through the end points of the contact area (Fig. 69·2). The circle pertaining to the maximum value of this shearing stress which, for  $(\phi_1 - \phi_2) = \pi/2$  is equal to  $p/\pi$ , has a diameter of the length of the contact areas. Assuming conditions of plane strain ( $e_{33} = 0$ ) the pressure  $p$  under which the first yielding occurs along this circle, which is the maximum elastic bearing pressure of the contact area, is obtained from the yield condition:

$$p_{\max \text{ El}} = \frac{\pi}{\sqrt{3}} s_0 = 1.81 s_0 \quad (69 \cdot 6)$$

Under pressures  $p > p_{\max \text{ El}}$  the plasticized area spreads and is contained between adjacent circles  $t_{\max} = \text{const}$ . However, the elastic solution given by eq. 69·3 can no longer be used since, with extending plastic region, the stresses in the elastic regions are changed.

For a pressure distribution  $p$  increasing parabolically towards the edges of the contact area, with a relation  $p_{\max} = 2p_0$ , the average maximum elastic bearing pressure exceeds by about 27 percent the pressure for uniform distribution; for a parabolic pressure distribution increasing from zero at the edge to a maximum at the center the bearing pressure is about 27 percent lower.<sup>68·4</sup>

### References

- 64·1 I. S. SOKOLNIKOFF, *Mathematical Theory of Elasticity*, McGraw-Hill Book Co., New York (1946) 121.
- 64·2 *Ibid.*, 187.
- 64·3 A. NADAI, *Z. angew. Math. & Mech.* **3** (1923) 442.
- 65·1 J. FRITSCHE, *Z. angew. Math. & Mech.* **11** (1931) 176.
- 65·2 A. NADAI, *Plasticity*, McGraw-Hill Book Co., New York (1931) 169.
- 65·3 F. B. SEELEY and J. O. SMITH, *Advanced Strength of Materials*, John Wiley & Sons, New York (1951).
- 65·4 A. THUM and F. WUNDERLICH, *Forschungsarb. Ingenieurwesen* **3** (1932) 261.
- 65·5 W. PRAGER, *Forschungsarb. Ingenieurwesen* **4** (1933) 95.
- 65·6 C. F. KOLLBRUNNER, *Pub. Intern. Assoc. Bridge & Structural Eng. Zurich*, **3** (1935) 222.

66·1 A. E. MACRAE, *Overstrain of Metals*, H. M. Stationery Office, London (1930).  
M. MALAVAL, *Rev. mét.* **20** (1923) 46; *Compt. rend.* **176** (1923) 448.  
A. NADAI, *op. cit.*, 186, and *Reissner Anniversary Volume*, J. W. Edwards Ann Arbor (1949) 430.

66·2 N. M. BELIAEV and A. K. SINITSKI, *Proc. Acad. Sci. USSR* (1938) 3.

66·3 S. TIMOSHENKO, *Theory of Elasticity*, McGraw-Hill Book Co., New York (1934) 323.

66·4 A. NADAI, *op. cit.*, 200.

67·1 H. HENCKY, *Z. angew Math. & Mech.* **4** (1924) 331.

67·2 A. NADAI, and L. H. DONELL, *Trans. ASME* **51** (1929) APM-51-16.

67·3 S. TIMOSHENKO, *op. cit.* **66**, 364.

67·4 P. G. HODGE and G. N. WHITE, *J. Applied Mech.* **17** (1950) A-5.

68·1 L. PRANDTL, *Z. angew Math. & Mech.* **1** (1921) 15.

68·2 A. NADAI, *ibid.*, 20.

68·3 G. SACHS, *Z. tech. Physik* **8** (1927) 132.

68·4 A. M. FREUDENTHAL, *Prelim. Rept. 2d Congr. Intern. Assoc. Bridge & Structural Engrs.*, W. Ernst, Berlin (1936) 10.

## PLASTICITY. PROBLEMS OF FLOW

**70. Two-Dimensional Problems. Glide Lines**

A state of two-dimensional stress within an ideal plastic material under conditions of slow steady flow is independent of the deformations; it can be determined from the equilibrium equations and the condition of plasticity. Since for plane strain ( $e_3 = 0$ ) the Huber-Mises-Hencky yield condition and the St. Venant condition differ only by a constant factor, the plasticity equation has the form,

$$(s_{11} - s_{22})^2 + 4s_{12}^2 = 4k^2 \quad (70 \cdot 1)$$

where  $k = s_0/\sqrt{3}$  or  $k = \frac{1}{2}s_0$ , respectively, depending on the yield condition applied. For plane stress ( $s_3 = 0$ ) the St. Venant condition is of the same form as for plane strain; the Huber-Mises-Hencky condition, however, takes the form,

$$s_{11}^2 + s_{22}^2 - s_{11}s_{22} + 3s_{12}^2 = s_0^2 \quad (70 \cdot 2)$$

Either of the eqs. 70·1 or 70·2, together with the equilibrium conditions, is sufficient to determine the stress components of the two-dimensional problem.

Solutions of the plane problem of plastic flow may be obtained by introducing a stress function  $F$ . If the stress components are expressed by

$$s_{11} = \frac{\partial^2 F}{\partial x_2^2}; \quad s_{22} = \frac{\partial^2 F}{\partial x_1^2}; \quad s_{12} = \frac{\partial^2 F}{\partial x \partial y} \quad (70 \cdot 3)$$

the relation between stress and strain and the compatibility

conditions lead in the elastic state to the well-known biharmonic differential equation of the Airy stress function. In the plastic state the relations of elasticity are replaced by the plasticity condition (eq. 70·1). Hence, the differential equation for the stress function of the plastic state,

$$\left( \frac{\partial^2 F}{\partial x_2^2} - \frac{\partial^2 F}{\partial x_1^2} \right)^2 + 4 \left( \frac{\partial^2 F}{\partial x_1 \partial x_2} \right)^2 = 4k^2 \quad (70 \cdot 4)$$

This is a nonlinear partial-differential equation of hyperbolic type, some special solutions of which have been given by Prandtl<sup>70·1</sup> and Nadai.<sup>70·2</sup> It possesses two families of characteristics, which are the orthogonal curves  $\alpha = \text{const}$ ,  $\beta = \text{const}$  in the  $(x_1 x_2)$  plane; they are one-parameter families of curves creating the integral surface of the differential equation. Oseen has shown<sup>70·3</sup> that, if these characteristics are introduced as a curvilinear system of coordinates, the nonlinear differential eq. 70·4 is transformed into two simultaneous linear hyperbolic differential equations with variable coefficients and the same characteristics. By a further transformation, of variables introducing the radii of curvature of the characteristics  $R_\alpha$  and  $R_\beta$  a linear hyperbolic differential equation with constant coefficients,

$$\frac{\partial^2 R}{\partial \alpha \partial \beta} + cR = 0 \quad (70 \cdot 5)$$

is obtained for either of the radii. For these equations integration methods have been developed.<sup>70·4</sup> Thus, solutions of stress problems of the fully plastic state are obtained, if the characteristics of the hyperbolic differential equations are known. It can be shown that at all points within the plastic region the directions of these characteristics are identical with the directions of the lines of principal shear; thus, the characteristics of the eq. 70·4 are the glide lines of the problem.

The use of the yield condition 70·2 in the solution of problems of plane stress leads to a nonlinear partial-differential equation of elliptic type for the stress function, of which no significant solutions are known.

In order to use glide lines for the solution of two-dimensional problems of plasticity it would be necessary to determine boundary values in respect to the radii of curvature of the glide lines

from the physical conditions of the problem, subsequently integrating the differential equations with regard to these boundary conditions. The applicability of the method thus depends essentially on the possibility of selecting adequate boundary conditions of the problem in terms of the radii of curvature of the glide lines.

The number of problems that have been solved by the method of glide lines is small. It is usually difficult to integrate the

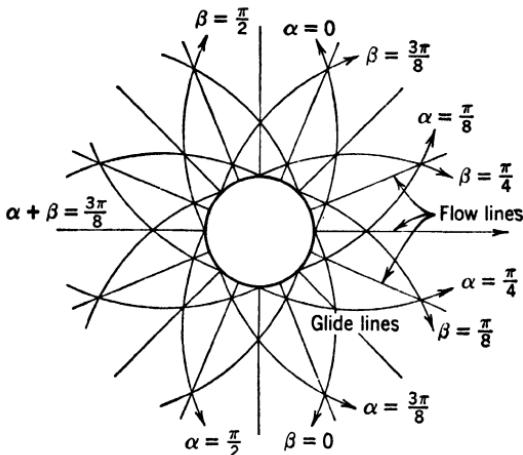


FIG. 70·1 Glide-line field of orthogonal logarithmic spirals (after Geiringer<sup>70·4</sup>).

partial-differential eqs. 70·5 directly for given boundary conditions. A more promising approach is therefore to select functions of which it is known that they are particular solutions of the differential eqs. 70·5 and to determine whether the boundary conditions they are able to satisfy have any physical reality.<sup>70·5</sup>

A simple function of this type is, for instance,

$$R_\alpha = R_\beta = C e^{(\alpha+\beta)} \quad (70\cdot6)$$

The pertaining field of glide lines  $\alpha = \text{const}$  and  $\beta = \text{const}$  consists of two families of orthogonal logarithmic spirals, which intersect on the radii through the origin, crossing them at angles of  $45^\circ$ . These radii and the concentric circles along which  $p = \text{const}$  are the lines of principal stresses of the plastic problem (Fig. 70·1), which is that of the circular hole under pressure in the infinite plane.

## 71. Technological Problems. Pressing and Rolling

Problems of plastic flow are encountered in the analysis of many technological processes involving the deformation of metals by pressing, rolling, and drawing. The knowledge of the forces developed and of the power required in such processes is of considerable practical interest. In first approximation these forces can be determined by theoretical analysis, considering the metal as an ideal plastic body. Solutions of the problems of pressing, rolling, and drawing have been obtained by v. Kármán,<sup>71·1</sup> Hencky,<sup>71·2</sup> Nadai,<sup>71·3</sup> Orowan,<sup>74·4</sup> and others.

**COMPRESSION OF A THIN SHEET.** The simplest problem is that of compression of a thin sheet of metal of thickness  $h$  between two parallel rigid plates. Under conditions of plane strain, the distribution of pressures  $p = f(x_1)$  along the parallel plates of length  $2l$  (Fig. 71·1) can be easily obtained, if in the yield condition 41·7 the influence of shearing stresses is neglected and the stress components  $s_{11}$ ,  $s_{22}$ , and  $s_{33}$  are taken as the principal stresses. Hence,

$$(s_{11} - s_{22})^2 + (s_{22} - s_{33})^2 + (s_{33} - s_{11})^2 \doteq 2s_0^2 \quad (71\cdot1)$$

Because of the assumption of a thin sheet the stresses over  $h$  may be assumed constant; hence:  $s_{22} = \text{const} = -p(x_1)$ . For  $e_3 = 0$ ,

$$s_{33} = \frac{1}{2}(s_{11} + s_{22}) = \frac{1}{2}(s_{11} - p) \quad (71\cdot2)$$

Eliminating  $s_{22}$  and  $s_{33}$  from the yield condition by introducing the relation 71·2, the simplified yield condition is obtained:

$$s_{11} + p = 2s_0/\sqrt{3} = \text{const} \quad (71\cdot3)$$

According to Fig. 71·1, the condition of equilibrium of the horizontal forces, considering the effect of a shearing stress  $t$  along the compression plates, is

$$\frac{ds_{11}}{dx_1} h = - \frac{dp}{dx_1} h = 2t \quad (71\cdot4)$$

If the shearing stress  $t$  is produced by dry friction  $t = \mu p$ , where  $\mu$  denotes the coefficient of friction, eq. 71·4 takes the form,

$$\frac{dp}{dx} = - \frac{2\mu p}{h} \quad (71\cdot5)$$

the solution of which in the range  $0 < x < l$  is given by

$$p = \frac{2s_0}{\sqrt{3}} e^{2\mu(l-x)/h} \quad (71 \cdot 6)$$

if the boundary conditions  $s_{11} = 0$  for  $x = +l$  and  $p = 2s_0/\sqrt{3}$  are to be fulfilled. Within the range  $-l < x < 0$ , the sign of  $x$  in the exponent of eq. 71·6 changes. The pressure distribution expressed by eq. 71·6 is represented in Fig. 71·1.

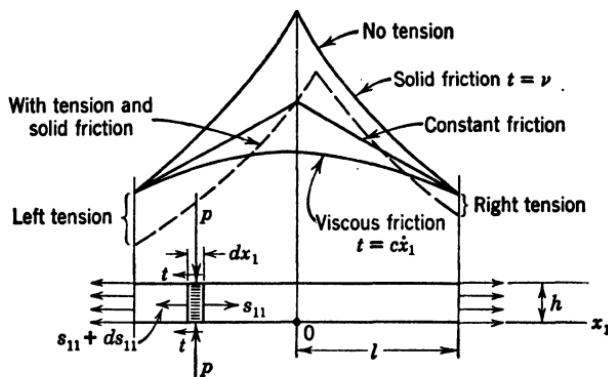


FIG. 71·1 Pressure distribution along compressed thin sheet (after Nadai<sup>71·3</sup>).

If the compressed sheet is subjected to a uniform tensile stress  $s$  at right angles to the direction of the compression, the pressure  $p$  follows the same exponential law but is reduced in proportion to  $\frac{(2s_0/\sqrt{3}) - s}{2s_0/\sqrt{3}}$ . Thus, for a tensile stress  $s_1$  acting at  $x = -l$ , the left part of the pressure curve becomes

$$p = \left( \frac{2s_0}{\sqrt{3}} - s_1 \right) e^{2\mu(l+x)/h} \quad (71 \cdot 7)$$

and, for  $s_2$  acting at  $x = +l$ , the right part of the pressure curve becomes

$$p = \left( \frac{2s_0}{\sqrt{3}} - s_2 \right) e^{2\mu(l+x)/h} \quad (71 \cdot 8)$$

These curves are shown in Fig. 71·1; their point of intersection defines a point in the compression plate at which the friction  $t$

changes its direction. Its position is variable and shifts away from the side on which a higher tensile stress  $s$  is applied.

If it is assumed that the shearing stress  $t$  is due to viscous resistance and therefore proportional to the relative velocity of slip between sheet and pressure plates, the shearing stress may, for symmetrical flow, be introduced as a linear function of  $x_1$ . Integrating eq. 71·4 with  $t = cx_1$  the pressure distribution becomes

$$p = \frac{2s_0}{\sqrt{3}} + (c/h)(l^2 - x^2) \quad (71 \cdot 9)$$

which is a parabolic distribution.

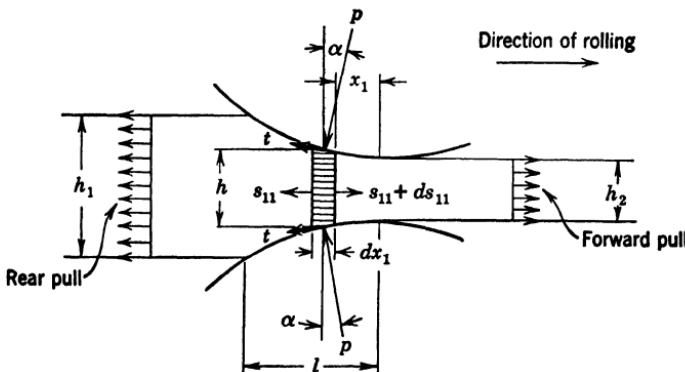


FIG. 71·2 Equilibrium condition of element in rolling of strip.

For a constant shear stress  $t = \text{const}$ , a linear distribution is obtained. Figure 71·1 shows the pressure distribution for the different assumptions concerning friction along the pressure plates.

**ROLLING OF A THIN STRIP.** The rolling pressure in a strip that is being simultaneously pulled through the rolls and the thickness of which is reduced from  $h_1$  to  $h_2$  can be computed by a similar method of approximation to that used previously. According to Fig. 71·2,  $r$  denotes the radii of the working rolls,  $h$  the variable thickness of the sheet in the contact region,  $x_1$  the horizontal distance of a point  $P$  in the sheet from the origin at the narrowest section between the rolls,  $l$  the length of the horizontal projection of the contact area, and  $\alpha$  the angle between the vertical and the

normal to the rolls at  $P$ . Denoting by

$$\lambda = \frac{h_1 - h_2}{h_1} \quad (71 \cdot 10)$$

the reduction ratio, and introducing the relations valid for small angles  $\alpha$

$$\frac{x_1}{r} = \sin \alpha \sim \alpha \quad \text{and} \quad (1 - \cos \alpha) = \frac{\alpha^2}{2} = \frac{x_1^2}{2r^2} \quad (71 \cdot 11)$$

the variable thickness of the sheet may be expressed by

$$h = h_2 - 2r(1 - \cos \alpha) = h_2 + \left( \frac{x_1^2}{r} \right) \quad (71 \cdot 12)$$

Since

$$r(1 - \cos \alpha) = \frac{1}{2}(h_1 - h_2) = \frac{l^2}{2r} \quad (71 \cdot 13)$$

the length of the contact area,

$$l = \sqrt{r(h_1 - h_2)} = \sqrt{\lambda r h_1} \quad (71 \cdot 14)$$

If  $p$  denotes the specific rolling pressure at the point  $P$ ,  $s_{11}$  the mean value of the stress in the sheet in the direction of rolling,  $t$  the friction between rolls and sheet at the point  $P$ , and  $v$  the mean horizontal velocity of the sheet at the point  $P$ , the equilibrium condition for the horizontal forces acting on a small prism of material of width  $dx_1$  and height  $h$  (Fig. 71·2) require that

$$\frac{d}{dx_1} (s_{11}h) + 2t + 2p \frac{x_1}{r} = 0 \quad (71 \cdot 15)$$

The yield condition is expressed by eq. 71·3. Because of volume constant flow,

$$vh = v_1 h_1 = v_2 h_2 \quad (71 \cdot 16)$$

Introducing eqs. 71·3 and 71·12 into 71·15, the differential equation of the rolling problem is obtained:

$$(x_1^2 + rh_2) \frac{dp}{dx_1} - \frac{4s_0}{\sqrt{3}} x_1 - 2t = 0 \quad (71 \cdot 17)$$

which has first been derived by v. Kármán.<sup>71·1</sup>

In order to integrate it, certain assumptions have to be made

concerning the shearing stress  $t$  which expresses the resistance due to the surface friction of the rolls. As in the previous example, these assumptions may be (a)  $t = \pm t_0$  (constant friction); (b)  $t = \pm \mu p$  (solid friction) and  $t = f(v)$  (viscous resistance). Nadai<sup>71-3</sup> has integrated eq. 71·17 for all three conditions and computed the distribution of the rolling pressure for different values of front and rear pull. For the simplest assumption  $t = \pm t_0$ , eq. 71·17 becomes

$$\frac{dp}{dx_1} = \frac{\pm 2t_0}{rh_2 + x_1^2} + \frac{4s_0}{\sqrt{3}} \cdot \frac{x_1}{rh_2 + x_1^2} \quad (71 \cdot 18)$$

which is transformed into

$$dp = \left[ \frac{\pm q}{1 + z^2} dz + \frac{2z}{1 + z^2} dz \right] \frac{2s_0}{\sqrt{3}} \quad (71 \cdot 19)$$

with  $q = \frac{t_0}{s_0} \sqrt{\frac{3}{rh_2}}$  and  $z = \frac{x_1}{\sqrt{rh_2}}$ . The solution of eq. 71·19 is

$$p = \frac{2s_0}{\sqrt{3}} [\pm q \tan^{-1} z + \log(1 + z^2)] + C \quad (71 \cdot 20)$$

The upper and lower signs of  $q$  are to be taken for forward slip and for backward slip, respectively, and the two integration constants  $C_1$  and  $C_2$  for the two signs determined from the conditions at the ends  $x_1 = 0$  and  $x_1 = l$  of the contact zone, so that the given front and rear tension  $s_{01}$  and  $s_{02}$  are attained there. Thus, from eq. 71·3,

$$s_{10} = \frac{2s_0}{\sqrt{3}} - p(x_1)_{(x_1=0)} = C_1 \text{ and } s_{20} = \frac{2s_0}{\sqrt{3}} - p(x_1)_{(x_1=l)} = C_2 \quad (71 \cdot 21)$$

or

$$C_1 = \frac{2s_0}{\sqrt{3}} - s_{10} \quad \text{and} \quad C_2 = \frac{2s_0}{\sqrt{3}} [1 + q \tan^{-1} l - \log(1 + l^2)] - s_{20} \quad (71 \cdot 22)$$

The point of intersection of the two pressure curves defined by the two signs of  $\pm q$  and the two constants  $C_1$  and  $C_2$ , respectively, determines the location of the change in the direction of the slip from forward to backward. The pressure peaks are the steeper,

the larger the friction; the intensity of the pressure is reduced by the application of front or rear tension, and the location of the pressure peaks shifts in the direction of the sheet movement if both tensions are equal. If one of the tensions is increased relative to the other, the location of the peak pressure is shifted towards the side of the smaller tension.

## 72. Technological Problems: Wire Drawing

The force required to draw a wire through a die depends essentially on the reduction of the wire area, the friction  $t$  of the wire on the walls of the die, and the yield condition of the wire. Introducing, as before, the simplifying conditions that

1. The axial, radial, and tangential stresses and strains are

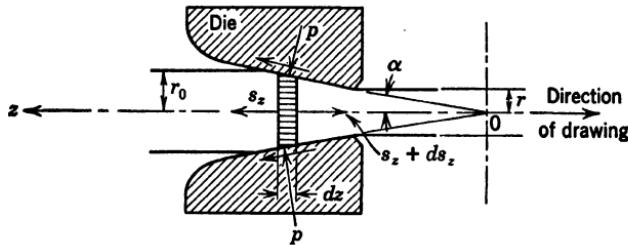


FIG. 72·1 Equilibrium condition of element in wire-drawing die.

principal strains, influence of the shear stress components being negligible.

2. The axial tension is uniformly distributed over the cross section; and

3. The radial stresses within the die are equal to the pressure  $p$  exerted by the die;

The condition of equilibrium in the direction of the drawing of a wire element passing through the die can be expressed by the equation (Fig. 72·1):

$$\frac{d}{dz} (s_z A) + 2tr\pi + 2pr\pi \sin \alpha = 0 \quad (72\cdot 1)$$

The strains in the wire are determined by the angle of the die. The radial and tangential strain increments are

$$de_\theta = \frac{dr}{r} \quad \text{and} \quad de_r = \frac{dr}{r} \quad (72\cdot 2)$$

Since for volume-constant deformation,

$$3de_v = de_z + de_r + de_\theta = 0 \quad (72 \cdot 3)$$

the strain component,

$$de_z = -2 \frac{dr}{r} \quad \text{and} \quad e_z = 2 \log \frac{r_0}{r} \quad (72 \cdot 4)$$

where  $r_0$  denotes the original wire radius; hence,

$$e_r = e_\theta = -\frac{e_z}{2} = \log \frac{r_0}{r} \quad (72 \cdot 5)$$

Thus

$$(e_\theta - e_r) = 0 \quad \text{and} \quad (e_r - e_z) = -(e_z - e_\theta) = 3 \log \left( \frac{r_0}{r} \right) \quad (72 \cdot 6)$$

Since

$$(s_r - s_z) = 2G(e_r - e_z) \quad \text{and} \quad (s_z - s_\theta) = 2G(e_z - e_\theta) \quad (72 \cdot 7)$$

the distortional energy yield condition in terms of stresses becomes

$$(s_r - s_z)^2 + (s_z - s_\theta)^2 = 72G^2 \left[ \log \left( \frac{r}{r_0} \right) \right]^2 = 2s_0^2 \quad (72 \cdot 8)$$

Because of eq. 72·5,  $s_r = s_\theta$ : moreover, it has been assumed that  $s_r = -p$ . Equation 72·8 may therefore be written in the form:

$$(s_z + p) = 6G \log \left( \frac{r}{r_0} \right) = s_0 \quad (72 \cdot 9)$$

Introducing solid friction  $t = \mu p$  and considering that  $z \cdot \sin \alpha = r$ , eq. 72·9 can be transformed into the differential equation,

$$r \frac{ds_z}{dr} + 2s_z + 2p \left( 1 + \frac{\mu}{\alpha} \right) = 0 \quad (72 \cdot 10)$$

where, for small die angles,  $\alpha$  has been introduced instead of  $\sin \alpha$ . Considering the relation 72·9, eq. 72·10 becomes

$$r \frac{ds_z}{dr} - 2ks_z + 2s_0(1 + k) = 0 \quad (72 \cdot 11)$$

where  $k = \mu/\alpha$ . The integral of this equation is

$$s_z = s_0 \frac{1+k}{k} \left[ 1 - \left( \frac{r}{r_0} \right)^{2k} \right] \quad (72 \cdot 12)$$

This expression has been developed by several investigators.<sup>72 \cdot 1</sup>

If the coefficient of friction  $\mu$  is assumed to be zero, eq. 72 \cdot 10 becomes

$$r \frac{ds_z}{dr} + 2s_z + 2p = 0 \quad (72 \cdot 13)$$

or, considering eq. 72 \cdot 9,

$$r \frac{ds_z}{dr} + 2s_0 = 0 \quad (72 \cdot 14)$$

the integral of which is

$$s_z = +2s_0 \log \left( \frac{r_0}{r} \right) \quad (72 \cdot 15)$$

The force necessary to draw a wire through a die, reducing its area from  $A_0$  to  $A$ , is therefore, according to eq. 72 \cdot 12,

$$P = s_z A = s_0 A \frac{1+k}{k} \left[ 1 - \left( \frac{A}{A_0} \right)^k \right] \quad (72 \cdot 16)$$

for  $\mu \neq 0$ , and according to eq. 72 \cdot 15,

$$P = s_z A = s_0 A \log \left( \frac{A_0}{A} \right) \quad (72 \cdot 17)$$

for vanishing die friction  $\mu = 0$ .

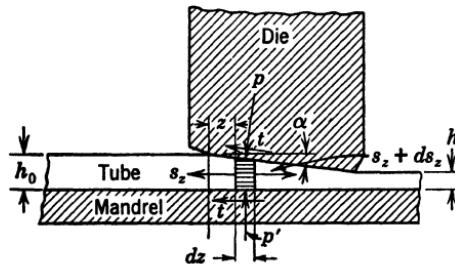


FIG. 72 \cdot 2 Equilibrium condition of element in drawing of thin tube.<sup>72 \cdot 2</sup>

Other problems, such as the drawing of a thin-walled tube (Fig. 72 \cdot 2) can be analyzed by the same approximate method,

considering the equilibrium conditions in the direction of the drawing of an element of width  $dz$ .<sup>72·2</sup>

The assumption of ideal plastic flow on which the foregoing analysis has been based does not adequately reproduce the behavior of most metals. In order to obtain results which would better reproduce the real behavior of metals at or near room temperature the effect of work hardening must be considered.

### References

- 70·1 L. PRANDTL, *Goettlinger Nachrichten* (1920); *Z. angew. Math. & Mech.* **6** (1923).
- 70·2 A. NADAI, *Z. Physik* **30** (1924) 106.
- 70·3 C. W. OSEEN, *Arkiv. Mat. Astron. och Fysik* **20** (1928); **24A** (1933).
- 70·4 H. GEIRINGER, *Mém. sci. math. Paris, fasc.* **86** (1937) 33.
- 70·5 A. NADAI, *Plasticity*, McGraw-Hill Book Co., New York (1931) 227.
- 71·1 TH.V. KÁRMÁN, *Z. angew. Math. & Mech.* **5** (1925) 139.
- 71·2 H. HENCKY, *Z. angew. Math. & Mech.* **5** (1925) 115.
- 71·3 A. NADAI, *Trans. ASME* **61** (1939) A-54.
- 71·4 E. OROWAN, *Proc. Inst. Mech. Engrs.* **150** (1944) 140.
- 72·1 E. A. DAVIS and S. J. DOKOS, *J. Applied Mechanics* **11** (1944) A-193.
- 72·2 G. SACHS, J. D. LUBAHN, and D. P. TRACY, *J. Applied Mechanics* **11** (1944) A-199.

## CHAPTER

## 15

## WORK HARDENING AND CREEP. SPECIAL PROBLEMS

## 73. Bending

WORK HARDENING. Analyzing the problem of volume-constant pure bending, which has been treated in Art. 65 for ideal plastic deformation, under the conditions of linear work hardening expressed by eq. 47·6, and retaining the Bernoulli-Navier

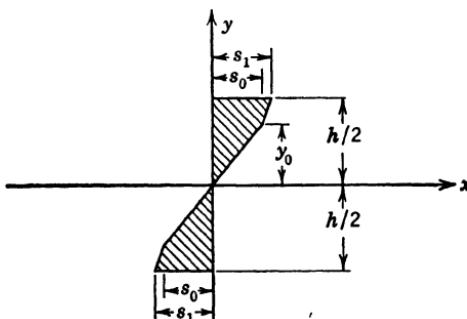


FIG. 73·1 Stress distribution in pure bending for linear work-hardening law.

assumption of a plane section remaining plane after deformation and the assumption of a uniaxial state of stress, the statical moment of the stresses with respect to the neutral axis in a section symmetrical to the neutral axis is given by eq. 65·1. The Bernoulli-Navier condition is expressed by equation 65·4; the difference in the analysis is introduced by the difference in the elastic-plastic distribution of stresses over the cross section

(Fig. 73·1). Introducing the relation 65·3 into eq. 65·4 gives

$$\Delta = \frac{h}{y} \cdot \frac{s_0}{3G} = 2e \quad (73\cdot1)$$

By substituting eq. 47·7 for  $e$ , the relation is obtained:

$$(s_1 - s_0) = \frac{H}{G} s_0 \left( \frac{h}{2y} - 1 \right) \quad (73\cdot2)$$

The statical moment of the stresses in the elastic-plastic state, according to eq. 65·1,

$$M_i = \int_{-h/2}^{+h/2} sy \, dA = 2 \int_0^{y_0} sy \, dA + 2s_0 \int_{y_0}^{h/2} y \, dA + 2 \int_{y_0}^{h/2} (s_1 - s_0)y \, dA \quad (73\cdot3)$$

Considering eq. 73·2 in the integration, the moment is obtained:

$$M_i = \frac{3}{2} M_{EI} \left[ 1 - \frac{1}{3} \left( \frac{\Delta_{EI}}{\Delta} \right)^2 + \frac{H}{G} \left( 1 - \frac{\Delta_{EI}}{\Delta} \right)^2 \right] = M \quad (73\cdot4)$$

Hence,

$$\left( \frac{\Delta_{EI}}{\Delta} \right)^2 - 3 \frac{H}{G} \left( 1 - \frac{\Delta_{EI}}{\Delta} \right)^2 = 3 - 2 \frac{M}{M_{EI}} \quad (73\cdot5)$$

or, because of eq. 65·4,

$$\eta^2 - 3 \frac{H}{G} (1 - \eta)^2 = 3 - 2 \frac{M(x)}{M_{EI}} \quad (73\cdot6)$$

where  $\eta = 2y_0/h$ . This is the equation from which the elastic-plastic boundary can be computed for any distribution of the external moment  $M(x)$ .

The shape of the deflection curve  $w(x)$  of the elastic-plastic part of the bent bar may be obtained from the relation,

$$\frac{d^2w}{dx^2} = - \frac{M(x)}{EI} \quad (73\cdot7)$$

considering eqs. 65·1, 65·4, and 65·3; the differential equation of the deflection,

$$\frac{d^2w}{dx^2} = - \frac{\Delta}{h} = - \frac{\Delta_{EI}}{2y_0} = - \frac{2s_0}{3Gh} \cdot \frac{1}{\eta(x)} \quad (73\cdot8)$$

**CREEP.** For linear materials the stress distribution is that of the elastic material (see Art. 35). Considering a nonlinear volume-constant viscous material the stress-steady creep rate relation of which is given by the power function,

$$\left(\frac{s}{s_1}\right) = \left(\frac{\dot{e}}{\dot{e}_1}\right)^n \quad (73 \cdot 9)$$

where  $s_1$  is the stress in the extreme fiber,  $n < 1$  and  $\dot{e}_1$  denotes the strain rate in the extreme fiber (Fig. 73·2) in conjunction with

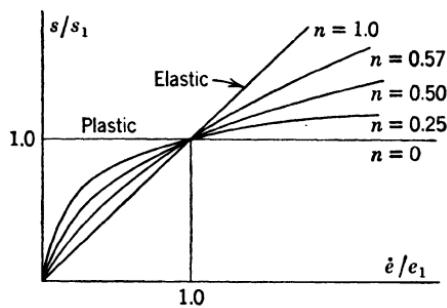


FIG. 73·2 Creep functions according to eq. 73·9 for different values of  $n$  (after Nadaï<sup>73·2</sup>).

the Bernoulli–Navier condition for plane sections in the viscous material,

$$\frac{\dot{e}}{\dot{e}_1} = \frac{2y}{h} \quad (73 \cdot 10)$$

the relation is obtained,

$$s = s_1 \left( \frac{2y}{h} \right)^n . \quad (73 \cdot 11)$$

Since for a symmetrical cross section the neutral axis is at the center of the section, the resisting moment,

$$M = 2 \int_0^{\frac{h}{2}} b s y \, dy \quad (73 \cdot 12)$$

where  $b(y)$  denotes the width of the section

With  $b = \text{const}$  and  $s$  from eq. 73·11,

$$M = \frac{2bs_1}{h^n} 2^n \int_0^{\frac{h}{2}} y^{(n+1)} dy = \frac{2}{n+2} bs_1 \left(\frac{h}{2}\right)^2 \quad (73\cdot13)$$

Hence, the extreme fiber stress,

$$s_1 = \frac{2(n+2)M}{bh^2} \quad (73\cdot14)$$

Thus, if  $M$  is kept constant during the test the steady-state stress distribution along a cross section remains constant (see Art. 35).

For  $n = 1$ , eq. 73·14 expresses the extreme fiber stress in an elastic beam; for  $n = 0$ , that in a fully plastic beam. The stress distribution is linear for  $n = 1$  and uniform for  $n = 0$ ; stress distributions for intermediate values  $0 < n < 1$  are curved according to eq. 73·11, as shown in Fig. 73·3.

The relation between extreme fiber strain and time is obtained by eliminating the stress  $s_1$  from eqs. 73·9 and 73·14:

$$\text{const } (\dot{e}_1)^n = \frac{2(n+2)M}{bh^2} = \text{const } M \quad (73\cdot15)$$

or

$$\frac{de}{dt} = \text{const } M^{\frac{1}{n}} \quad (73\cdot16)$$

Hence, the strain-time curve has the form: \*

$$e = \text{const } M^{\frac{1}{n}} t \quad (73\cdot17)$$

The preceding equations have been derived for steady-state creep in which the stress distribution is stable and the creep rates in the fibers are proportional to their respective distances from the neutral axis, as required by eq. 73·10. However, before this state is reached, a transient stage must be passed as the initially elastic *linear* distribution of stress, which constitutes the

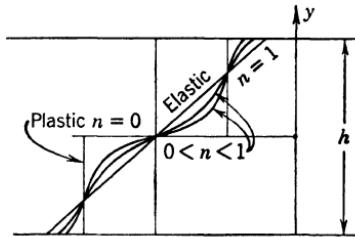


FIG. 73·3 Stress distribution in pure bending according to eq. 73·11 for values of  $0 < n < 1$ .

instantaneous response of the section at the time  $t = 0$  and is associated with the condition that the instantaneous strains in the fibers are proportional to their respective distances from the axis, is transformed into the steady-state *nonlinear* distribution. This transient stage is usually of no practical significance; its length depends obviously on the creep characteristics of the material as expressed, for instance, by the power in eq. 73·9; the less  $n$  differs from one the shorter the time necessary to arrive at the steady-state stress distribution.

#### 74. Infinitely Long Cylindrical Hole under Pressure within Infinite Plane

**WORK HARDENING.** According to eqs. 26·5 and 47·15, introducing the notation  $\bar{H} = 2H(1 + \mu)$  and  $\bar{G} = 2G(1 + \mu)$ , the compatibility relations in cylindrical coordinates are

$$\begin{aligned}\bar{\epsilon}_r &= \frac{\partial u_r}{\partial r} = \frac{1}{\bar{H}} (s_r - \mu s_\theta - s_{0r} + \mu s_{0\theta}) + \frac{1}{\bar{G}} (s_{0r} - \mu s_{0\theta}) \\ \bar{\epsilon}_\theta &= \frac{u_r}{r} = \frac{1}{\bar{H}} (s_\theta - \mu s_r - s_{0\theta} + \mu s_{0r}) + \frac{1}{\bar{G}} (s_{0\theta} - \mu s_{0r}) \quad (74\cdot 1) \\ \bar{\epsilon}_z &= \frac{du_z}{dz} = 0\end{aligned}$$

Eliminating  $s_\theta$  from the eqs. 74·1 by substituting the expression for  $s_\theta$  from the equilibrium condition 26·2 and subsequently differentiating the second of the eqs. 74·1 with respect to  $r$  and comparing it with the first, the differential equation for  $s_r$  is obtained in the form:<sup>74·1</sup>

$$r^2 \frac{d^2 s_r}{dr^2} + 3r \frac{ds_r}{dr} = \frac{s_0}{\sqrt{3}} 2(1 + \mu) \left( 1 - \frac{\bar{H}}{\bar{G}} \right) \quad (74\cdot 2)$$

The solution for the elastic plate with a circular hole is given by eq. 66·46,

$$s_{re} = -s_{\theta e} = \frac{s_0}{\sqrt{3}} \left( \frac{r_0}{r} \right)^2$$

where  $r_0$  denotes the radius of the elastic-plastic boundary. For  $r = r_0$  the stresses  $s_{r0}$  and  $s_{\theta 0}$  satisfy the Huber-Mises-Hencky yield condition. The boundary conditions along  $r = r_0$  which

the solution of eq. 74·2 must satisfy are:  $s_{re} = s_r = s_{0r}$ ;  $s_{\theta e} = s_\theta = s_{0\theta}$ ;  $u_{re} = u_r$ ; moreover, for  $r = r_i$  the internal pressure  $q = -s_r$ . The solution of eq. 74·2 under these boundary conditions is given by the function,

$$s_r = \frac{s_0}{\sqrt{3}} \left\{ (1 + \mu) \left( 1 - \frac{\bar{H}}{\bar{G}} \right) \left[ \left( \frac{r_0}{r} \right)^2 + 2 \log \left( \frac{r}{r_0} \right) - 1 \right] - \left( \frac{r_0}{r} \right)^2 \right\} \quad (74 \cdot 3)$$

and, from eq. 74·1,

$$s_\theta = \frac{s_0}{\sqrt{3}} \left\{ (1 + \mu) \left( 1 - \frac{\bar{H}}{\bar{G}} \right) \left[ 1 - \left( \frac{r_0}{r} \right)^2 + 2 \log \left( \frac{r}{r_0} \right) \right] + \left( \frac{r_0}{r} \right)^2 \right\} \quad (74 \cdot 4)$$

For  $\bar{H} = 0$  and  $\mu = 0$  eqs. 74·3 and 74·4 are identical with the first two eqs. 66·45.

CREEP. Introducing the general solution 66·14 of the differential eq. 66·13 for strains under conditions of rotational symmetry and volume-constant deformation into the compatibility conditions 26·5, the strain-velocity components,

$$-\dot{e}_r = \dot{e}_\theta = \frac{\text{const}}{r^2} \quad (74 \cdot 5)$$

The intensities of stress and of strain velocity (designated here by the subscript  $i$  instead of  $r$  to avoid confusion with radial components),

$$s_i = \frac{\sqrt{3}}{2} (s_\theta - s_r) \quad \text{and} \quad \dot{e}_i = \frac{2}{\sqrt{3}} \dot{e}_\theta = \frac{\text{const}}{r^2} \quad (74 \cdot 6)$$

if  $s_z = \frac{1}{2}(s_\theta + s_r)$ . Hence, if for a nonlinear material a general power law of the type 73·9, expressed in terms of intensities of stress and strain,

$$s_i = \text{const } \dot{e}_i^n \quad (74 \cdot 7)$$

is assumed to govern the deformational behavior, the relation is obtained,

$$(s_\theta - s_r) = \text{const } r^{-2n} \quad (74 \cdot 8)$$

The equilibrium conditions 26·2 become

$$\frac{ds_r}{dr} = \frac{s_\theta - s_r}{r} = \text{const } r^{-2n-1} \quad (74\cdot 9)$$

which is the differential equation for the radial stress. Integration of eq. 74·9 gives the stress equations of rotational symmetry:

$$\begin{aligned} s_r &= C_1 + C_2 r^{-2n} \\ s_\theta &= C_1 + (1 - 2n)C_2 r^{-2n} \end{aligned} \quad (74\cdot 10)$$

By satisfying the boundary conditions  $s_r = 0$  for  $r = \infty$  and  $s_r = -q_i$  for  $r = r_i$ , the expressions are obtained,

$$\begin{aligned} s_r &= q_i \rho^{-2n} \\ s_\theta &= q_i(1 - 2n)\rho^{-2n} \end{aligned} \quad (74\cdot 11)$$

where  $\rho = r/r_i$ . By introducing  $n = 1$  the elastic or linear viscous distribution of stress (eq. 66·46) is obtained; the stress distribution within the ideal plastic material (eq. 66·45) is obtained by direct integration of eq. 74·9 for  $n = 0$ . For  $n = 0.5$  the tangential stress vanishes.

For the thick-walled cylinder Nadai has given the solution,<sup>74·2</sup>

$$\begin{aligned} s_r &= c[1 - \rho^{-2n}] \\ s_\theta &= c[1 - (1 - 2n)\rho^{-2n}] \end{aligned} \quad (74\cdot 12)$$

where  $\rho = r/r_e$ ;  $\rho_i = r_i/r_e$ , and  $c = q_i \rho_i^{2n} / (1 - \rho_i^{2n})$ , and has shown that, for values of  $0 < n < 1$ , the distribution of tangential stresses which reaches its maximum value at the inner surface of an elastic cylinder ( $n = 1$ ) gradually shifts outward and, for exponents  $n < 0.5$ , attains its maximum at the outer surface of the wall. For  $n = 0.5$ , a uniform distribution of the tangential stresses over the wall,

$$s_\theta = q_i \rho_i / (1 - \rho_i) \quad (74\cdot 13)$$

is obtained. It may be inferred from eqs. 74·11 that the distribution of radial stresses is considerably less sensitive to a variation of the exponent  $n$  than that of tangential stresses.

The creep strain is obtained from eqs. 74·6 and 74·7; both  $e_\theta$  and  $e_r$  are necessarily functions of the form  $f(n)\rho^{-2}$ .

## 75. Wire Drawing

As an example of a technological problem the process of wire drawing dealt with in Art. 72 under assumption of plastic flow

will be analyzed for linear work hardening. Because of eq. 72·5 and the resulting condition  $s_r = s_\theta$ , the stress and strain intensities, according to eqs. 41·12 and 41·13, are

$$s_i = s_r - s_z; \quad e_i = \frac{2}{3}(e_r - e_z) \quad (75\cdot1)$$

Hence, eq. 47·6 becomes

$$\frac{1}{2H} [(s_r - s_z) - (s_{r0} - s_{z0})] = (e_r - e_z) - (e_{r0} - e_{z0}) \quad (75\cdot2)$$

where  $s_{r0}$  and  $s_{z0}$  must fulfill the yield condition, which for  $s_{r0} = s_{\theta 0}$  has the form,

$$s_{r0} - s_{z0} = s_0 \quad (75\cdot3)$$

Neglecting in eq. 75·2 the elastic against the plastic strain and introducing eqs. 75·3 and 72·8, the work-hardening condition is transformed into the equation:

$$s_z - s_r - s_0 + 6H \log \left( \frac{r}{r_0} \right) = 0 \quad (75\cdot4)$$

Substituting eq. 75·4 and  $s_r = -p$  into the differential eq. 72·11 the new differential equation for wire drawing with work hardening is obtained,

$$r \frac{ds_z}{dr} - 2ks_z + 2(1+k) \left[ s_0 - 6H \log \left( \frac{r}{r_0} \right) \right] = 0 \quad (75\cdot5)$$

the solution of which is

$$s_z = \frac{1+k}{k} s_0 - \frac{6(1+k)H}{k} \log \left( \frac{r}{r_0} \right) - \frac{6(1+k)H}{k^2} + Cr^{2k} \quad (75\cdot6)$$

Determining  $C$  from the condition that at the die entrance  $r = r_0$  the stress  $s_z = 0$ , the force required to draw the wire through the die,

$$P = s_z A = \frac{1+k}{k} \left[ 3H \log \frac{A_0}{A} - \left( \frac{3H}{k} - s_0 \right) \left[ 1 - \left( \frac{A}{A_0} \right)^k \right] A \right] \quad (75\cdot7)$$

For  $H = 0$  this equation is identical with eq. 72·16.

### References

74·1 K. SWAINGER, *Phil. Mag.* (7)36 (1945) 454.  
74·2 A. NADAI, *J. Applied Phys.* 8 (1937) 424.

## CHAPTER

# 16

## DESIGN FOR PLASTICITY AND WORK HARDENING

### 76. Limitations of Design for Elasticity

The design of an engineering structure or part of it can generally be reduced to the consideration of two problems:

1. The *determination* of the stresses in and the deformations of the structure resulting from the external loads.
2. The *evaluation* of the significance of those stresses and deformations in terms of the capacity of the structure or of the material of which it is composed, for carrying these loads.

In the conventional design which is mostly a design for elasticity, the design procedure is based on the assumptions that:

- (a) Materials are perfectly elastic and stresses are proportional to loads;
- (b) The capacity of the material and of the structure to carry a load is exhausted when the computed maximum stress attains the level at which the material ceases to respond elastically under the considered conditions of stress, unless fracture by separation (fatigue), or failure by instability (buckling) occurs at a lower stress level; in that case this latter limit defines the load-carrying capacity of the structure.

Although methods of structural analysis and design are based on these two assumptions, it is taken for granted that the actual structural performance is not governed by them. If it were, and if these assumptions were more than an expedient attempt to simplify the two basic problems of design, no structure could be

designed that would be "safe," because of the sharp stress concentrations that are present in every real structure. This fact is clearly recognized by all designers in the design of structural connections, which, contrary to the conventional design for elasticity of structural members, is actually a design for plasticity. No assembly, riveted, welded, or bolted, could practically be designed for elastic performance, that is, without the assumption of uniform distribution over the elements of the assembly of the response to the applied force, which implies that the material is able to relieve its peak stresses by yielding. Thus, in fact, the real strength of a material is less in its capacity to resist than in its ability to yield—not without limit since the performance of engineering structures depends on their retaining their given initial shape within the relatively narrow limits of tolerance—but by the right amount and at the appropriate rate. The inelasticity of the material determines therefore its structural performance considerably more than the separation strength.

Hence, linearity of the stress-strain diagram of a material is a desirable property only so far as it creates the theoretical conditions for the application of the relatively simple methods of classical elastic theory. It is the very slight deviation from elasticity, being the expression of the ability of the material to relieve, by limited yielding, localized excessive stresses, that creates the practical conditions for the application of elastic theory in the design of engineering structures.

The recognition of the difference of the approach to design under different conditions must be accompanied by an appraisal of all relevant facts. For instance, it is usually assumed that the possibility of design for plasticity as expressed by the assumption of relief of stress concentrations in metals depends on the shape of the stress-strain curve determined by a conventional test; it is however not sufficiently realized that such correlation exists only if the strain rate and the state of stress are the same in the test and under service conditions, or if the shape of the stress-strain diagram is not appreciably affected by the strain rate. This consideration necessarily leads to a different evaluation of the effects of stress concentrations and of the possibility of stress relief for basically different types of inelasticity.

If engineering structures and parts were designed for the sole function of carrying loads, a theoretically balanced design would

provide a uniform resistance of all parts and elements to the applied load; the specified limiting condition of resistance would then be attained simultaneously in all parts and elements of the structure. However, since the load-carrying capacity is frequently not the primary but a secondary (no matter how vital) consideration, whereas the primary function or some other consideration determines the shape of the structure, the possibilities of attaining a balanced design of uniform resistance are limited to structures the shape of which is primarily determined by the functions of load carrying, and the practical design of which is not too much affected by considerations concerning the desirable uniformity of elements used in the construction. Where the shape of the structure or of the part is unrelated to its load-carrying capacity or where, for economic reasons, the geometrical uniformity of the constituent elements is more important than the uniformity of their resistance, this resistance will attain a limiting value at certain critical points or sections, before such value is reached in the remaining parts of the structure. The significance of the limiting resistance value must then be considered in order to evaluate the significance of its being attained at certain individual points or sections only.

A limiting resistance value may be specified in three different ways:

1. In terms of a critical value of stress which defines failure by separation, occurring either immediately on the attainment of this value or after a period the length of which is a function of this critical stress.
2. In terms of a condition of instability.
3. In terms of a limiting, recoverable, or irrecoverable deformation or strain which defines extreme conditions of functional adequacy.

Conditions 1 and 2 represent *structural damage*; condition 3 represents *functional damage*.

If the limiting condition is specified in terms of a separation or fracture stress, it is usually assumed that the load-carrying capacity of the whole structure is seriously impaired when this stress is reached at one or a number of localized points or sections. Since the propagation of a crack, once it has been locally

formed, is a relatively rapid process, its duration should not be taken into account in design; the critical structural damage has already been done at the moment of initiation of the crack. Similarly, the specification of a limiting condition of instability of a critical section or member usually implies that the load-carrying capacity of the entire structure is endangered by this instability, although in some phases of modern aircraft design (thin webs and skin) local instability is not considered to represent a condition of failure of the whole structure. This latter approach is, however, the exception and is limited to secondary members buckling within the elastic range; it is not applied in the design of principal load-resisting parts.

If the limiting resistance is specified in terms of a deformation, the value of which is derived from considerations of the proper functioning of the structure in other respects than its mechanical resistance to loads, the question of structural resistance or damage does not arise. Damage is then defined in functional terms only, the structural adequacy under the limiting conditions being taken for granted.

The specification of limiting conditions in terms of either fracture or instability is equivalent to the application of the *weakest-link* concept used in the analysis of brittle fracture (see Art. 55), since it is assumed that the limit of resistance of the weakest or the most highly stressed member or part determines the limit of resistance of the structure. This approach is justified if the immediate result of the destruction of the weakest link is catastrophic, or if the resistance of the whole structure with regard to future service is so much affected by the failing of this link that a condition of near failure is attained, as in design for brittle fracture. However, in the conventional design for elasticity under steady or slowly varying loads, the *weakest-link* concept is retained, while the limiting condition in terms of brittle fracture (structural damage) is tacitly replaced by a limiting condition in terms of deformation (functional damage) represented by the yield limit, without considering that for such a limiting condition the *weakest-link* concept is no longer applicable. In design for plasticity the *weakest-link* concept is replaced by the concept of redistribution of the response to the applied load over the members or sections of the structure as a result of their

inelastic deformation, and the limiting condition in terms of functional damage by a condition of instability of the entire structure, which is again a true condition of structural damage.

The difference between these two approaches to design will be the more pronounced, the less the distribution of forces or bending moments follows the distribution of elastic resistance, the more extensive therefore the redistribution of the response to the applied forces that is possible and that leads to a stable equilibrium beyond the limiting conditions of conventional (elastic) design. The evaluation of this difference forms the subject of the theory of design for plasticity, for which the term *limit design* is frequently used. It should, however, be kept in mind that the specific use of this designation is hardly justified, since every rational design is necessarily a limit design; it is the specification of the limiting condition that may differ for different approaches to design.

Some problems of design for plasticity have been dealt with in Chapter 13 where the theory of plasticity has been applied to problems of elastic-plastic equilibrium and to the determination of conditions delimiting stable states of contained deformation and unstable states of free flow; these latter represent limiting conditions in design for plasticity. Thus, the critical internal pressure  $q_c$  of a thick-walled cylinder is not defined by the pressure at which yielding first develops at the inner surface, but by a considerably higher value, that is, by the pressure at which the yielding reaches the outside of the wall and the cylinder starts to flow freely. Similarly, the critical carrying capacity  $M_{Fl}$  of sections in torsion and in bending considerably exceeds the elastic carrying capacities  $M_{el}$ .

The principles on which design for plasticity must be based have been developed in the analysis of the structural model shown in Fig. 43·1 and have been illustrated by Figs. 43·2 and 43·3. In the practical application of those principles, in the design of metal structures, it is necessary to consider the consequence of abandoning the conventional assumption that the stresses in the structure are proportional to the applied loads; this assumption is valid only for perfect linear elasticity. The nonlinearity of the load-stress relation within the range of elastic-plastic deformation eliminates the linearity of the relation between the stresses under service conditions which are usually

elastic and the stresses associated with the limiting resistance. Therefore the conventional definition of a *factor of safety* as the ratio between the two stress levels, which makes the use of the concept of *allowable* or *working stresses* possible, loses its meaning, since the stress ratio is no longer identical with the load ratio, in terms of which alone safety can rationally be defined. Hence, in design for plasticity the fundamental definition of the safety factor as the ratio between the ultimate carrying capacity of the structure and the service load should be applied instead of the safety factor expressed in terms of a stress ratio. The two definitions are equivalent only in design for elasticity, that is, for fracture or for instability.

### 77. The Factor of Safety

The conventional concept of *allowable stress* implies a comparison between a computed maximum stress under the acting loads and the fracture stress of the material; it also implies the existence of a margin between the two stresses. As its conventional name *margin of safety* suggests, it reveals the subjective striving on the part of the designer for an adequate measure of safety as well as a consciousness of the limitations of his knowledge and the arbitrariness of his assumptions. The real character of the margin of safety has remained obscure, however, and its magnitude is generally estimated on the basis of subjective judgment rather than objective fact. The most refined design is thus deprived of its merits, since the designer is free to select the fundamental assumption of his design largely on the basis of subjective arguments, without being compelled to ascertain their validity by the identification of the objective conditions. Research in the sphere of new materials cannot be expected to bear its full weight on the economy of structures if the safety factor can thus be fixed rather arbitrarily.

The principle underlying the concept of the safety factor can best be understood by reviewing the fundamental difficulty of structural design. The computed structural characteristic (*stress*) cannot be equated to the characteristic derived from observable and measurable physical properties (*resistance*), because the results of an intellectual process (design) cannot be equated to the result of a material perception (test result). The limitations of human observation are such that no observable

quality can be measured *exactly*. The designer is able to assert only that a value is *larger than* a lower limit and *smaller than* an upper limit—not that it is *equal* to another value. The manifest lack of correspondence between the conceived and the performed action in any sphere of human activity explains why no real identity may be expected in a series of actions or events that were planned to be identical. Consequently, intellectual concepts, contrived to reproduce material phenomena with a certain grade of perfection, may be correlated with material observations regarding those phenomena only by a relation of *inequality*, expressed in terms of a probable correlation range. This range itself will be a function of the degree of perfection in the concept. It must therefore provide for: (1) the imperfection of human observations and actions (uncertainty), and (2) the inadequacy of intellectual concepts devised to reproduce physical phenomena (ignorance). This range represents the objective minimum value of the safety margin, which is thus identified as a function of objective uncertainty as well as of subjective ignorance.

With increasing perfection of design methods, the element of *ignorance* is largely eliminated; the element of uncertainty however is caused by circumstances that can, to a certain extent, be changed but that can never be removed. Hence, the safety factor is a measure of uncertainty rather than of ignorance. The trend toward reducing its numerical value is not so much the result of improved design methods as it is the result of modified objective circumstances; that is, of standardizing engineering materials by introducing quality control in production, of applying standard acceptance tests by the users of such materials, and of introducing stringent regulations for the control of workmanship.

The laws of structural design are derived from the principles of classical mechanics and are based on the existence of a causal relationship between the antecedent and the consequent events. They are mostly expressed in the form of differential equations, the solution of which enables the engineer to determine all the consequences following one or a number of given antecedent events. Therefore, within the range for which the initial assumptions are valid, the designer should be able to go confidently from cause to effect, all phenomena concerned being strictly predictable.

A certain number of parameters of these equations however represent observable and measurable physical properties or phenomena. The application of the differential equations to structural design requires the introduction of the real values of such properties under all conceivable conditions of practical importance. Some of these values must be predicted or estimated on the basis of past experience, since their observation and measurement under all relevant conditions is impracticable. Such prediction is entirely different from that based on differential equations since the effect of *chance* inherent in any prediction from past experience becomes significant. The concept of deterministic causality is thus superseded by a new concept, in which every unknown cause is termed a *chance cause*. Systems of chance causes produce events in accordance with the law of large numbers and thus give rise to statistical laws represented by frequency distributions. Therefore prediction based on statistical causality can be expressed only in terms of the *probability* of a certain event to occur. Hence, an observable physical property can be represented only by a frequency distribution of the observed values so that the constancy of any physical property is of a purely statistical nature. A given quality approaches a constant value only in the sense that it may be represented by a genuine frequency distribution produced by chance causes. The laws of structural design, therefore, must be considered a combination of functional and statistical relationships, functional so far as the laws of the theory of structures are concerned, and statistical to the extent that real physical properties appear as parameters of the functional relations.

The value of the safety factor  $m$  may be derived from the condition that the maximum service load  $P_s$  (or moment) which the structure has to carry must never cause such damage as to impede the fitness for service of the structure, even if the maximum service load were to coincide with the lowest possible carrying capacity  $P_r$ . Hence,

$$m = \frac{P_r}{P_s} > 1 \quad (77 \cdot 1)$$

If  $(\pm \Delta P_s)$  denotes the maximum range of fluctuation of the service load about its expected or most probable value  $P_{s0}$ , and  $(\pm \Delta P_r)$  the maximum range of fluctuation of the carrying capac-

ity of the structure about its most probable value  $P_{r0}$ , the maximum service load will be  $(P_{s0} + \Delta P_{s0})$  and the minimum carrying capacity will never be less than  $(P_{r0} - \Delta P_r)$ . According to eq. 77·1, failure is prevented if

$$P_{s0} + \Delta P_s < P_{r0} - \Delta P_r \quad (77\cdot 2)$$

Therefore,

$$P_{s0} < \frac{1 - \frac{\Delta P_r}{P_{r0}}}{1 + \frac{\Delta P_s}{P_{s0}}} P_{r0} = \frac{1}{m} P_{r0} \quad (77\cdot 3)$$

The value of the factor of safety  $m$  depends on the variation of the parameters of the design (loading conditions, dimensions, weights, mechanical properties); the ranges of fluctuation of the individual parameters determine the ultimate ranges of variation of load and of carrying capacity in accordance with the law of statistical superposition. A structure is thus designed by predicting its future behavior on the strength of knowledge gained by past experience. The computation of its safety factor requires an analysis of the variability of all influences bearing on its resistance and on the service loads and conditions.<sup>77·1</sup>

The factor of safety is thus affected by two groups of influences: (a) Influences that govern the stress induced in the structure or the load that produces such stress; and (b) Influences that govern the resistance of the structure or its carrying capacity.

The general approach to the design, as expressed by the selection of the limiting conditions, determines the *mechanism of resistance*, that is, the assumed mode of computation of the ultimate carrying capacity. Since it is usually impossible to conceive such a mechanism that will effectively reproduce the actual behavior of the structure, but which, at the same time, is simple enough to be suitable for practical design, every devised mechanism is fictitious to a certain degree. But if it embodies relevant physical properties, which are determinable by relatively simple standardized tests, and contains a reasonable theoretical concept, correlating the principal design parameters in a dimensionally correct form, its practical efficiency and reliability can be ascertained by experiments, the statistical interpretation of which will furnish information concerning the range of dispersion

of individual values about the line of best fit. The suitability of the conceived mechanism is usually judged by the simplicity of the underlying concept, by the closeness with which experimental results are reproduced, and by the narrowness of the range of dispersion of such results about the line of best fit which represents the average trend.

In the design of metal parts or structures the ultimate carrying capacity is represented by the load that produces free plastic flow, whereas the elastic carrying capacity is represented by the load that produces the first contained local yielding. It is doubtful whether the limit of free isotropic flow can actually be attained; it has been observed that failure by instability occurs under loads that are appreciably lower than the theoretically computed limits (see Art. 80). This reduction of the ultimate load is apparently due to the heterogeneity and anisotropy of the spreading plastic deformation, which produces conditions approximating flow by the spreading of glide lines through the elastic domain under a load that is smaller than that necessary to produce isotropic flow. It appears therefore that neither design for elasticity nor design for full isotropic plasticity constitute an adequate procedure. It would, however, be justified to consider the elastic carrying capacity and the load defining full plastic flow as the two extremes, enclosing the possible range of carrying capacities.

Whereas in design for elasticity the limiting condition for a single load application and for repeated loading is identical unless the problem of fatigue enters into the considerations, these two limits are different in design for plasticity. The difference between the two load levels depends on the intensity of the residual stresses set up in the structure or part by the first load application (see Art. 43).

## 78. Design for Plasticity of Redundant Engineering Structures (Theory of Limit Design)

In the design for plasticity of plane or spatial structural parts the difference between the limiting condition for plastic and for elastic design is determined by the nonuniformity of the distribution of stress or of elastic resistance. In the case of linear statically indeterminate engineering structures, it is the degree of redundancy and the distribution of such characteristics of

resistance as moments of inertia or sectional areas that determine this difference. The total difference is made up of the difference between the limiting conditions associated with individual critical sections or members and the difference between the limiting conditions for the entire structure due to its redundancy. In statically determinate structures the second component does not exist; the first component only, with regard to the critically stressed section or member, determines the difference, which, according to eq. 65·6, may be considerable for rectangular sections in bending, though not for the conventional thin-webbed flanged sections.

The redundancy of a structure the resistance of which is well balanced with respect to a certain loading condition would by itself not provide a carrying capacity exceeding that associated with design for elasticity. However, since variable loading has usually to be considered, no balanced distribution of resistance is practically possible with respect to any individual loading condition, so that an excess of the elastic-plastic carrying capacity over the elastic capacity may be expected to exist in any redundant structure.

The carrying capacity of a  $n$ -fold redundant structure subject to bending moments resulting from the single application of a load is defined by the load producing a statically determinate state of unrestrained flow, in which all  $n$  redundants and one nonredundant section have attained their limiting values. These values depend on the yield point and the dimensions of the cross sections. Thus, an  $n$ -fold redundant structure under increasing load is transformed into a structure of gradually decreasing degree of redundancy, before failing finally as a statically determinate structure under a load at which the stable elastic-plastic equilibrium is transformed into an unstable state of free plastic flow. The gradual reduction in redundancy of an  $n$ -fold redundant structure is associated with the formation of *plastic hinges* at the  $n$  sections, at which the bending moments  $M$  consecutively exceed the limiting elastic moment  $M_{El}$ , and gradually tend towards the limiting plastic moment of the section  $M_{Pl}$ . Prior to failure the state of the redundant structure is determined by the system of  $n$  redundant moments  $X_k = M_{kPl}$  acting at the plastic hinges;  $M_{kPl}$  denotes the limiting value of the plastic moment at the redundant section  $k$ .

Under a load exceeding that which produces this determinate condition, the resistance of  $n$  sections no longer increases; the sections act like plastic hinges with a resisting capacity limited to the moment  $M_{kpl}$ . The formation of the  $(n + 1)$ th plastic hinge at the section of maximum moment of the determinate structure transforms it into an unstable structure. Hence, the limiting condition, defined by a limit of stress in the design for elasticity, is transformed into a stability limit in the design for plasticity.

However, this is only the case if the deformation of the material is ideally plastic. In a work-hardening material the gradual reduction of the degree of redundancy does not take place; the limiting condition is therefore not automatically determined by a condition of instability, but has to be designated by the introduction of an additional criterion, either in terms of a deformation specified as excessive, or in terms of a maximum stress leading to fracture. In this difference in the character of the limiting conditions is the principal difference in the approach to the design for plasticity and for work hardening.

Since in structures of work-hardening materials the redundants do not reach limiting values, but beyond the elastic limit increase steadily with increasing load at a rate determined by the work-hardening coefficient  $H$  (see Art. 47), it is difficult to determine the relation between the redundants of the system and the load under increasing load intensity. Redundant structures of linear work-hardening materials therefore have to be analyzed step by step, as the linear elastic relation between stress and strain defined by the modulus  $G$  gradually changes into a linear work-hardening relation defined by the modulus  $H$ , and the linear relation between load and stress is transformed into a complex nonlinear one. Because of this changing behavior, the simultaneous equations for the redundants also become nonlinear.

The carrying capacity under repeatedly applied or variable loading of an  $n$ -fold redundant structure subject to bending moments is defined by the maximum load, the moments of which, in combination with the residual moments resulting from the previous application of any other load, produce stresses that at no point exceed the yield limit of the material. Hence, if  $n$  states of inherent (load-independent) moments are defined by

selecting the redundants so that for each state one redundant  $X_k = 1$ , while all the others are zero, and if  $\sigma_{ik}$  denotes the maximum stress at the section  $i$  resulting from the state of inherent moments produced in the statically determinate basic structure by  $X_k = 1$  (while all other redundants are zero) the carrying capacity of the redundant structure under repeated load is defined by the maximum load fulfilling the conditions,<sup>78.1</sup>

$$\left| s_i + \sum_{k=1}^{k=n} \sigma_{ik} X_k \right| \leq s_0 \quad \text{and} \quad \left| \sum_{k=1}^{k=n} \sigma_{ik} X_k \right| \leq s_0 \quad (78.1)$$

where  $s_i$  denotes the stress that would be produced by the limiting load in the statically indeterminate structure of unlimited elasticity, and  $X_k$  are arbitrarily selected values of the redundants. Condition 78.2 can be extended to cover combinations of various loads if it is written in the form,

$$\begin{aligned} & \left| \max s_i + \sum_{k=1}^{k=n} \sigma_{ik} X_k \right| \leq s_0 \\ & \left| \min s_i + \sum_{k=1}^{k=n} \sigma_{ik} X_k \right| \leq s_0 \\ & \left| s_{iw} + \sum_{k=1}^{k=n} \sigma_{ik} X_k \right| \leq s_0 \end{aligned} \quad (78.2)$$

where  $|\max s_i|$  and  $|\min s_i|$  denote respectively the maximum positive and negative values of the stresses at the section  $i$  produced in the statically indeterminate structure of unlimited elasticity by any combination of loads, including dead load;  $s_{iw}$  denotes the dead load stress at  $i$ ; and  $X_k$  arbitrarily selected values of the redundants. Thus, if for an  $n$ -fold statically indeterminate structure subject to a specified load in variable position it is possible to select a system of redundants  $X_1 \cdots X_k \cdots X_n$  so, that the conditions 78.2 are fulfilled, the structure is able to carry this load in any position, as well as in any possible combination of alternating positions.

The eqs. 78.2 imply that if in a redundant structure a set of redundants can be selected so as to define a condition of the structure that would represent the mobilization by plasticity

of the possible maximum of *self-help*, the structure will actually tend to attain that condition. This process has been termed "shake-down" by Prager.<sup>78-2</sup>

The application of the principles of design for plasticity to the design of statically indeterminate *trusses* is limited, because of the fact that a member the carrying capacity of which is reached in compression cannot be considered in the same light as one that fails in tension, since the force-deformation relationship beyond the limiting load is different in both cases. In the case

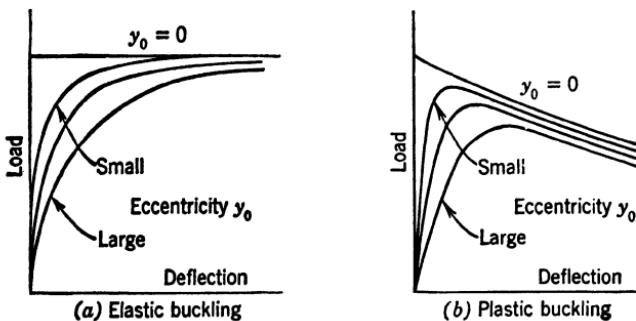


FIG. 78-1 Force-deformation curves for elastic and plastic buckling (after v. Kármán<sup>81-2</sup>).

of axial tension the limiting load of the member or section is reached when the uniform stress has attained the yield limit. Beyond this limit the deformation increases indefinitely under the practically constant limiting force. If the member represents a redundant member of a statically indeterminate structure, the limiting value of the redundant force is therefore constant and independent of the deformation. In the case of slender compression members the force-deformation relationship characteristic for buckling produces a different behavior in the elastic and in the plastic range. Figure 78-1 shows that in the elastic range the buckling load is constant and practically independent of deformation; in the plastic range, however, the force, after having reached the stability limit, drops appreciably with increasing lateral deflection of the member. Therefore, the carrying capacity of compression members under conditions of plastic buckling decreases rapidly, with the result that the difference between the ultimate and the momentary forces has to be carried by other members of the structure.

Considering the behavior of the structural model represented in Fig. 43·1 subject to a compression force  $P$ , the carrying capacity may be reached under different conditions, which depend on the relative dimensions of the members 1 and 2 and on the angle  $\alpha$ :

(a) The buckling limit of the bar  $S_2$  is attained before the stress  $s_2$  reaches the yield limit. Under increasing lateral deflection and very nearly constant buckling resistance of  $S_2$ , the forces in  $S_1$  increase gradually with increasing load until their buckling load and thus the capacity load of the entire structure is attained. In this case the limiting load is determined by the sum of the limiting (elastic buckling) loads of the individual bars, in the same manner as in tension.

(b) The buckling limit of the bar  $S_2$  is reached after the stress  $s_2$  has reached the yield limit. As indicated in Fig. 78·1, the resistance of  $S_2$  decreases considerably with increasing lateral deflection, with the result that the forces in  $S_1$  increase rapidly with increase of the load  $P$  until they attain the buckling load, and the structure fails. In this case the limiting load of the structure is determined by the buckling load of the bars  $S_1$  only, since the force in  $S_2$ , after appreciable lateral deflection of the member, drops to a fraction of its buckling load.<sup>78·3</sup>

Since the slenderness ratios of bars that fail by buckling in the elastic range are so large that the forces they are able to carry are usually small enough to be neglected in a truss the other members of which do not fail by buckling in the elastic range, and since bars failing by buckling in the plastic range cannot be relied on to carry permanently any load, members that fail in buckling should not be considered at all, if methods of design for plasticity are applied to statically indeterminate trusses. It is therefore possible to apply this method only if none of the redundant members, the carrying capacity of which is reached before the limiting load of the truss is attained, fails in buckling.

In such trusses, a limiting state can gradually be reached under repeatedly applied load and defined by the condition that all deformations finally become elastic as a result of the built-up system of residual stresses; the maximum load for which this state is possible is the ultimate carrying capacity. The following

conditions under which it can be reached have been derived by Gruening:<sup>78-4</sup>

1. If under the applied load the stresses in  $n$  members of an  $n$ -fold redundant truss, which can be considered the redundants of the basic determinate system, exceed the yield limit, they will be reduced to stresses below this limit under load repetition, if the stresses in one of the members of the basic system exceed the yield limit.

2. If the stresses in  $n$  redundant members and in a few members of the basic system exceed the yield limit, those stresses will be reduced to values below this limit under repeated application of the load, if the forces in the yielding members of the basic system and the forces in the redundant members are connected by such conditions of equilibrium that both decrease simultaneously.

3. If the stresses in a single member of the basic system, which is connected with the redundant members by such conditions of equilibrium that a reduction of the forces in any one of the redundant members is accompanied by an increase of those stresses, exceed the yield limit before the stresses in all redundant members have been reduced to values below this limit by repeatedly applied load, the total deformation produced by the repeatedly applied load tends to increase indefinitely, and the stresses remain permanently above the yield limit.

In the analysis of problems of contained elastic-plastic deformation it is usually assumed that the principle of minimum potential energy governs the elastic-plastic equilibrium in the same way as it governs the elastic equilibrium, the only difference being the condition of plasticity which, in the elastic-plastic problem, enters the problem of variation of the elastic potential  $\delta\Phi = 0$  as an auxiliary condition (see Art. 43). If the validity of this consideration is extended to statically indeterminate structures, it follows that the evaluation of the redundants of an indeterminate structure under elastic-plastic conditions can be based on the same variational principles that are applied in the theory of elastic structures. Hence, if  $\Phi$  denotes the stored-up elastic strain energy of the structure expressed in stresses, forces, or moments, the successive variation with regard to the redundants  $X_k$  leads to the  $k$  conditions from which the values of  $X_k$  can be determined:

$$\frac{\delta\Phi}{\delta X_k} = y_k = 0 \quad (78 \cdot 3)$$

where  $y_k$  denotes the deformation in the direction of the redundant  $X_k$ . The difficulty in the application of this generalized principle of Castiglano is in the evaluation of the function  $\Phi$  separately for the different elastic and elastic-plastic portions of the structure, the individual length of which depend in turn on the values of the redundants themselves. The eqs. 78·1 are nonlinear and can, even for rather simple structures, only be solved by methods of approximation.

In the case of elastic deformation the applied strain energy

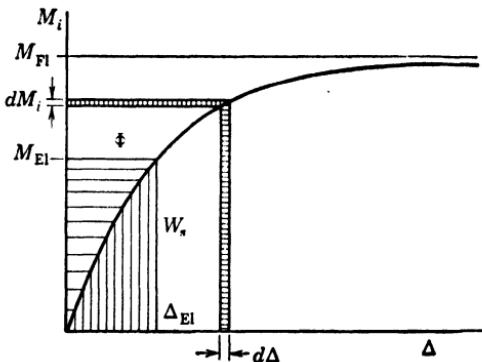


FIG. 78·2 Relation between bending moment and sum of extreme fiber strains  $\Delta$ .

$W_s = \Phi$  and the dissipated energy  $W_d = 0$ ; for elastic-plastic deformation  $W_s = \Phi + W_d$ . Considering eq. 65·2 and the relation between the bending moment  $M_i$  and the sum of the extreme fiber strains  $\Delta$  in the elastic-plastic state represented in Fig. 78·2, the stored-up potential energy is expressed by<sup>78·5</sup>

$$d\Phi = \left[ \int_0^{M_{EI}} dM_i \left( \frac{\Delta}{h} \right) + \int_{M_{EI}}^M dM_i \left( \frac{\Delta}{h} \right) \right] dx \quad (78 \cdot 4)$$

while the total strain energy,

$$dW_s = \left[ \int_0^{\Delta_{EI}} M_i d\left(\frac{\Delta}{h}\right) + \int_{\Delta_{EI}}^{\Delta} M_i d\left(\frac{\Delta}{h}\right) \right] dx \quad (78 \cdot 5)$$

Only within the elastic range  $\Delta \leq \Delta_{EI}$  and  $M \leq M_{EI}$  is  $\int dW_s = \int d\Phi_s$ , as may be seen by comparing the areas above and below the  $M_i - \Delta$  diagram in Fig. 78·2.

Expressing  $d\Phi$  in terms of deformation,

$$d\Phi = \left\{ \int_0^{\Delta_{EI}} [M - M_i(\Delta)] d\left(\frac{\Delta}{h}\right) + \int_{\Delta_{EI}}^{\Delta} [M - M_i(\Delta)] d\left(\frac{\Delta}{h}\right) \right\} dx \quad (78 \cdot 6)$$

and introducing for the rectangular section  $M_i$  according to eq. 65·6, as well as replacing  $M_i$  by  $M$  after the integration has been performed, the relation is obtained:

$$d\Phi = \frac{1}{2} \cdot \frac{M_{EI}^2}{EI} \left( 3 - 2 \frac{\Delta_{EI}}{\Delta} \right) dx \quad (78 \cdot 7)$$

Similarly,

$$dW_s = \frac{1}{2} \cdot \frac{M_{EI}^2}{EI} \left( \frac{\Delta_{EI}}{\Delta} + 3 \frac{\Delta}{\Delta_{EI}} - 3 \right) dx \quad (78 \cdot 8)$$

The dissipated energy,

$$dW_D = dW_s - d\Phi = \frac{3}{2} \cdot \frac{M_{EI}^2}{EI} \cdot \frac{\Delta}{\Delta_{EI}} \left( 1 - \frac{\Delta_{EI}}{\Delta} \right)^2 dx \quad (78 \cdot 9)$$

In terms of moments, introducing eq. 65·7, the potential energy,

$$d\Phi = \frac{1}{2} \cdot \frac{M_{EI}^2}{EI} \left( 3 - 2 \sqrt{3 - \frac{2M}{M_{EI}}} dx \right) \quad (78 \cdot 10)$$

Whereas for elastic deformation the potential energy increases indefinitely with increasing moment or deformation, this energy tends towards a limit in the elastic-plastic state, which is defined in Fig. 78·2 by the horizontal asymptote  $M = M_{Pl}$ . The dissipated energy,

$$dW_D = \frac{3M_{EI}^2}{EI} \left( \frac{2M_{EI} - M}{M_{EI} \sqrt{3 - \frac{2M}{M_{EI}}}} \right) dx \quad (78 \cdot 11)$$

This expression is positive only for  $M_{Pl} > M > M_{EI}$ .

## 79. Procedures of Limit Design

Procedures of limit design are based on the principle that the final values of the redundants may be selected arbitrarily so as to produce the most favorable distribution of moments or of forces.

The practical procedure is different in the design for a single application of the load, for repeated application of the same load, and for repetition of different loads. A different procedure is necessary if it is required to determine not only the final carrying capacity, but also the real values of the redundants of the elastic-plastic system as functions of the load. Finally, consideration of the real form of the stress-strain curve or of a linear work-hardening relation instead of the ideal plastic stress-strain relation requires another modification of procedure.

**SINGLE LOAD APPLICATION.** The carrying capacity under a single application of a load can be determined without analyzing

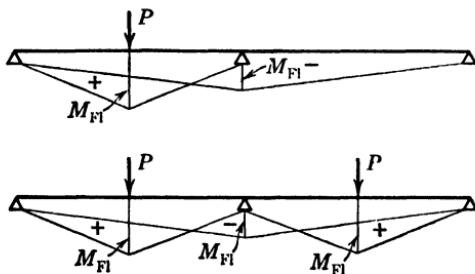


FIG. 79-1 Limiting condition for two-span beam.

the elastic redundant structure, by attributing to all redundant moments values equal to the ultimate values of the bending moment  $M_{fl}$  which the respective redundant sections and the critical section of the nonredundant structure are able to carry under the assumption of fully plastic distribution of stress. Hence, the final moment distribution of the continuous beam of two spans with uniform cross section (Fig. 79-1) is determined by the bending moments over the support and under the load becoming equal to  $M_{fl}$ . The ultimate load  $P$  is therefore obtained from the relation:

$$\frac{P_{fl}l}{4} = \frac{3}{2} M_{fl} \quad \text{or} \quad P_{fl} = \frac{6M_{fl}}{l} \quad (79-1)$$

The sequence of the formation of the plastic hinges does not affect the final state. Since the carrying capacity of the simply supported beam  $P_{0fl} = 4M_{fl}/l$ , the ratio of the carrying capaci-

ties for the simply supported and the continuous beam of two spans is 1:1.5. On the basis of design for elasticity ( $M_m = \frac{13}{16}Pl$ ) this ratio is,

$$\frac{4M_{EI}}{l} : \frac{64M_{EI}}{13l} = 1:1.23 \quad (79 \cdot 2)$$

Under the action of concentrated loads  $P$  at the centers of both spans, the carrying capacity for design for plasticity remains unchanged whereas that for elasticity (critical moment over support) becomes  $P = \frac{16}{3} \cdot \frac{M_{EI}}{l}$ ; hence, the ratio,

$$\frac{4M_{EI}}{l} : \frac{16M_{EI}}{3l} = 1:1.33 \quad (79 \cdot 3)$$

A moderate yielding of the central support does not affect the ultimate carrying capacity; it may only change the sequence of the formation of plastic hinges. Under two symmetrical concentrated loads or a uniformly distributed load the plastic deformation of the two-span beam starts over the central support. A small yielding of this support reduces the negative support moment sufficiently to transfer the first plastic deformation into the spans. As a result of the bending moment  $M > M_{EI}$  over a certain length near the center of the spans, the apparent rigidity of this length is reduced; the whole structure becomes less susceptible to further yielding. After plastic hinges have formed at mid-span the structure is no longer affected by yielding of the central support. Hence, it is only the sequence of the elimination of the redundants, which has been changed by the yielding, not the carrying capacity. In design for plasticity a moderate movement of supports need therefore not be considered.

**REPEATED APPLICATION OF THE SAME LOAD.** In order to determine the carrying capacity under repeated application of the same load, both the plastic and the elastic moment distribution of the redundant structure must be known in order to find the residual moments introduced by the plastic deformation. The ultimate distribution of moments under a single application of the limiting load for the continuous beam of two spans as previously determined, and the elastic distributions are as follows:  
For a concentrated load at mid-span of first span:

$$\text{Plastic: } P_{\text{fl}} = \frac{6M_{\text{fl}}}{l}; \quad M_m = -X = M_{\text{fl}}$$

$$\text{Elastic: } P = P_{\text{fl}}; \quad M_m = \frac{1}{8}P_{\text{fl}}l; \quad X = -\frac{3}{16}P_{\text{fl}}l \quad (79 \cdot 4)$$

For concentrated loads at mid-span of both spans:

$$\text{Plastic: } P_{\text{fl}} = \frac{6M_{\text{fl}}}{l}; \quad M_m = -X = M_{\text{fl}};$$

$$\text{Elastic: } P = P_{\text{fl}}; \quad M_m = \frac{5}{32}P_{\text{fl}}l; \quad X = -\frac{3}{16}P_{\text{fl}}l \quad (79 \cdot 5)$$

If the ultimate load is applied the first time and released, the residual moments are obtained by subtracting from the moments of the limiting plastic distributions the moments produced by the force ( $-P_{\text{fl}}$ ) in an elastic structure (see Art. 43). Hence, for a concentrated load at mid-span of the first span:

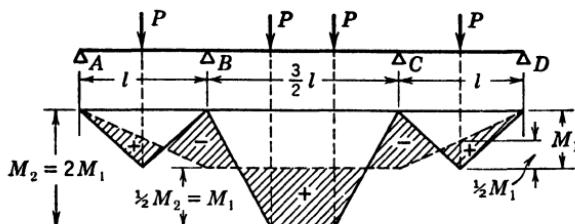
$$M_{m\text{Res}} = -\frac{7}{16}P_{\text{fl}}l; \quad X_{\text{Res}} = -\frac{1}{16}P_{\text{fl}}l; \quad (79 \cdot 6)$$

For concentrated loads at mid-span of both spans:

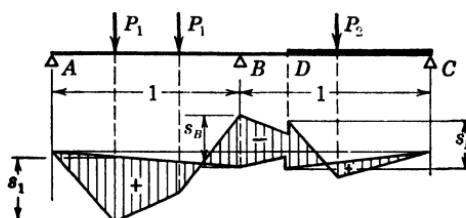
$$M_{m\text{Res}} = +\frac{2}{16}P_{\text{fl}}l; \quad X_{\text{Res}} = +\frac{4}{16}P_{\text{fl}}l \quad (79 \cdot 7)$$

Evidently, these residual moments appear only if the structure after unloading remains a continuous beam; otherwise, as a result of the plastic deformation, the beam will bend upward and rise from its central support. The full load  $P_{\text{fl}}$  can be repeated a second time only if no plastic deformation occurs on unloading and during the second loading. Hence, the sum of the maximum residual stresses in the critical sections and the elastic stresses produced in those sections by the residual moments must not exceed the yield limit. The maximum residual stress in sections in which the fully plastic stress distribution has been reached remains smaller than  $\frac{1}{2}s_0$  for all structural shapes, attaining the value  $\frac{1}{2}s_0$  for the full rectangular section only. Conditions where plastic deformation may occur on unloading are thus limited to diamond-shaped and round sections, where residual stresses may be higher than  $\frac{1}{2}s_0$ , since the elastic fiber stresses due to the residual moments will usually be considerably less than  $\frac{1}{2}s_0$ . If  $ks_0$  denotes the residual stress of the critical section in terms of the yield limit, the elasticity of the unloading process requires that

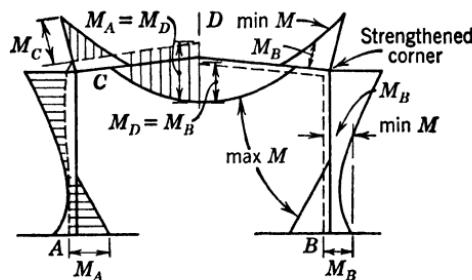
$$ks_0 + \frac{M_{\text{Res}}}{S} \leq s_0 \quad \text{or} \quad M_{\text{Res}} \leq s_0S(1 - k) \quad (79 \cdot 8)$$



(a) Continuous beam over three spans with line of support moments selected so that  $M_B = M_C = -\frac{1}{2}M_2$ .



(b) Continuous beam over two spans with discontinuous change of cross section at  $D$ . Line of support moment selected so that  $s_1 = s_B$  and  $s_D < s_B$ .



(c) Rigid frame under the action of vertical load and wind. Line of corner moments selected so that  $M_A = M_C = M_{D1}$  (left side) or for strengthening corner so that the resisting moment outside the corner  $< M_B$  (right side).

FIG. 79-2 Distribution of moments or stresses in redundant structures under steady loads according to principles of design for plasticity (limit design) (after Bleich<sup>78-1</sup>).

where  $S$  denotes the elastic section modulus. Only for small differences  $(1 - k)$  will this condition lead to limit loads  $P < P_{fl}$ .

**ALTERNATING APPLICATION OF DIFFERENT LOADS OR MOBILE LOADS.** If it is assumed that the two-span continuous beam is alternately subject to (1) one load  $P$  at mid-span of one span, and (2) symmetrical loads  $P$  at mid-span of both spans, the identity at the critical sections of the signs of the moment pertaining to one loading condition and the residual moment of the other condition implies that for either condition  $P < P_{fl}$ , if recurrent plastic deformation is to be avoided. The load  $P$  which can be sustained, if alternately applied in the two positions, is obtained from the conditions that the sum of the stresses produced at mid-span by the loading (1) and the residual stresses from the preceding loading (2), as well as the sum of the stresses produced over the support by the loading (2) and the residual stresses from the loading (1) do not exceed the yield stress  $s_0$ .

Instead of the detailed analysis of either condition, eqs. 78·2 may be applied to alternating and to mobile loads. With

$$s_{imax} = \frac{13}{64} \cdot \frac{Pl}{S}; \quad s_{imin} = -\frac{12}{64} \cdot \frac{Pl}{S}$$

and

$$\sigma_m = -\frac{1}{2}; \quad \sigma_{support} = 1$$

the residual moment is obtained from the condition,

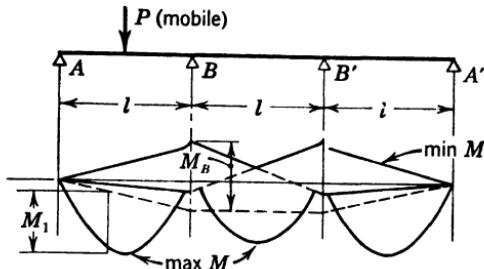
$$\frac{13}{64} \cdot \frac{Pl}{S} - \frac{1}{2} \cdot \frac{X}{S} = \frac{12}{64} \cdot \frac{Pl}{S} + \frac{X}{S} = s_0$$

Hence

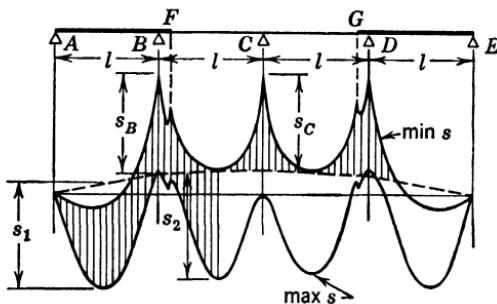
$$X = \frac{1}{96} Pl \quad \text{and} \quad P = \frac{96}{19} \cdot \frac{M_{fl}}{l} < P_{fl}$$

The general procedure based on eqs. 78·1 and 78·2 to be adopted for any type of loading, including mobile loads, is illustrated in Fig. 79·2 and 79·3. For a definite load the elastic distribution of moments and stresses must be established first. If the cross section is uniform, *moment* distributions may be used directly. If changes of cross section occur, maximum *stresses* must be computed; their distribution over the structure will show discontinuities at the points of changes of cross section. For mobile loads, elastic maximum and minimum moments and

the resulting stresses are computed for a number of sections by appropriately loading the moment-influence lines of the respective sections. After tracing the enveloping lines of maximum and minimum bending moments, the lines of the redundants (dash lines) are constructed in accordance with the conditions 78·2.



(a) Continuous beam over three spans subject to mobile load  $P$ , with line of support moments selected so that  $M_1 = M_B$ .



(b) Continuous beam over four spans with discontinuous change of cross section at  $F$  and  $G$  subject to uniform dead load and live load in critical position. Line of support moments selected so that  $s_1 = s_B = s_2 = s_C$ .

FIG. 79·3 Distribution of maximum moments or stresses in redundant structures under mobile loads according to principles of design for plasticity (limit design) (after Bleich<sup>78 1)</sup>.

**ANALYSIS OF REDUNDANT ELASTIC-PLASTIC SYSTEM.** The redundants of a structure which, over a certain length, has reached the elastic-plastic condition  $M > M_{el}$ , may be computed by the generalized energy variation method if the potential energy  $\Phi$  is determined separately for the elastic and for the elastic-plastic range. Since the length of the elastic-plastic section varies with the load, the analysis must be repeated for

different loads and for different values of the redundants  $X$ . The abscissas associated with the minimum ordinates of the curves  $\Phi = f(X)$  indicate the values  $X$  for every load.

For structures with cross sections for which the difference between  $M_{EI}$  and  $M_{FI}$  is relatively small, the variational procedure may be applied to  $\Phi$  obtained from the elastic structure with the equation of plasticity  $M = M_{EI}$  as an auxiliary condition.

**CONSIDERATION OF THE ACTUAL FORM OF THE STRESS-STRAIN DIAGRAM.** For redundant structures made of metals that do not show a sharp yield limit and for which therefore the stress-strain diagram of ideal plasticity does not represent a satisfactory approximation, the approach to design for plasticity is necessarily different, since the ultimate carrying capacity, instead of being automatically defined as a limit of stability, must be specified arbitrarily by introducing a separate criterion. The assumption of the equalization of moments by the formation of plastic hinges is not justified, and the structure remains fully redundant, even if the plastic deformation in all redundant sections extends over the whole depth of the cross sections. Since the length and the apparent rigidity of the elastic-plastic parts of the structure vary with the load, the relations between the redundants and the load intensity is nonlinear, without however, tending towards a final value determined by conditions of unrestrained plastic flow.

The computation is elementary, but the solution of the system of nonlinear equations by methods of approximation is extremely cumbersome, even for the simplest structures and loading conditions.

Prager has investigated the effect of different forms of the transition between elastic and plastic conditions on the behavior of a simple redundant structure.<sup>79-1</sup> For a general curved transition, the analysis of a fixed-end beam with symmetrical loads  $P$  in the third points showed that the shape of the transition curve did not appreciably affect the value of the redundant moment  $M$ . It was however of considerable influence on the *deformation* of the structure under load, particularly under loads approaching the ultimate carrying capacity.

## 80. Theory of Limit Design and Experiments

Experiments on redundant structures, particularly continuous beams, have shown that in metals with a sharp yield limit the

carrying capacity associated with fully plastic distribution of stress over the critical sections represents an ultimate theoretical limit and that it is a rational procedure to consider both this limit and the elastic carrying capacity as extremes, enclosing the possible range of fluctuations of the real carrying capacity, the expected value of which is in the vicinity of the mid-point between the extremes (see Art. 77).

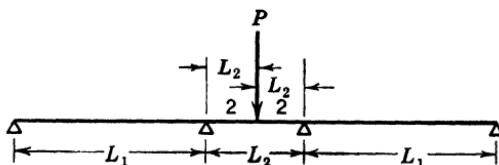


FIG. 80·1 Continuous beam for tests of elastic-plastic performance.

If this procedure is applied to both the individual cross sections and to the whole structure represented, together with the load, in Fig. 80·1, the limiting moment of the individual section,

$$M_{L0} = \frac{1}{2}(M_{El} + M_{Pl}) = \frac{1}{2}M_{El} \left(1 + \frac{T}{S}\right) \quad (80\cdot1)$$

where  $T$  denotes the plastic and  $S$  the elastic section modulus. If the ratio,

$$\alpha = \frac{3L_2}{4L_1 + 6L_2} \quad (80\cdot2)$$

the elastic carrying capacity,

$$P_{El} = \frac{4S_m s_0}{L_2(1 - \alpha)} \quad (80\cdot3)$$

where  $S_m$  denotes the elastic section modulus at mid-span of the central span. For a freely supported central span  $L_1 = \infty$  and  $\alpha = 0$ . Hence, for such a span,

$$P_0 = \frac{4S_m s_0}{L_2} \quad (80\cdot4)$$

and

$$\frac{P_{El}}{P_0} = \frac{1}{1 - \alpha} \quad (80\cdot5)$$

This ratio varies between  $P_{\text{El}}/P_0 = 1$  for a freely supported span and  $P_{\text{El}}/P_0 = 2$  for a central span with fixed ends defined by  $L_1 = 0$  and  $\alpha = 0.5$ .

The full plastic capacity load  $P_{\text{Pl}}$  is reached with the formation of two plastic hinges over the supports and one at mid-span of the central span. It is therefore independent of  $\alpha$  and can be expressed by

$$P_{\text{Pl}} = \frac{4(T_m + T_s)S_0}{L_2} \quad (80 \cdot 6)$$

where  $T_m$  and  $T_s$  denote the plastic section moduli at mid-span and over the supports, respectively. For constant section along the structure  $T_m = T_s = T$ , and

$$P_{\text{Pl}} = \frac{8Ts_0}{L_2} \quad (80 \cdot 7)$$

so that

$$\frac{P_{\text{Pl}}}{P_0} = \frac{2T}{S} \quad (80 \cdot 8)$$

Under the assumption that the expected value of the carrying capacity is the mid-point value,

$$P_{L0} = \frac{1}{2}(P_{\text{El}} + P_{\text{Pl}}) \quad (80 \cdot 9)$$

the relations are obtained,

$$P_{L0} = \frac{2M_{L0}}{L_2} \cdot \frac{3 - 2\alpha}{1 - \alpha} \quad (80 \cdot 10)$$

and

$$\frac{P_{L0}}{P_0} = \frac{1}{4} \left(1 + \frac{T}{S}\right) \frac{3 - 2\alpha}{1 - \alpha} \quad (80 \cdot 11)$$

In Fig. 80·2 eqs. 80·5, 80·8, and 80·11 are compared with results of tests by Stuessi, Kollbrunner, and Maier-Leibnitz.<sup>80·1</sup> The experimental values are scattered within a relatively narrow range about the mid-point values computed from eq. 80·11 and are thus considerably below the theoretical maximum (eq. 80·8).

Equation 80·11 becomes invalid for values of  $\alpha$  approaching zero, since the stability of the continuous beam with very large side spans and a plastic hinge at mid-span of the central span

vanishes rapidly as  $\alpha$  approaches zero. For  $\alpha = 0$  the carrying capacity is therefore given by  $P = \frac{1}{2}P_{EI}(1 + T/S)$ ; near  $\alpha = 0$  it falls off sharply towards this value from that given by eq. 80·11. This example is thus also an illustration of a limitation implicit in the theory of plastic redistribution of moments of  $n$ -fold redundant structures: the method of redistribution can

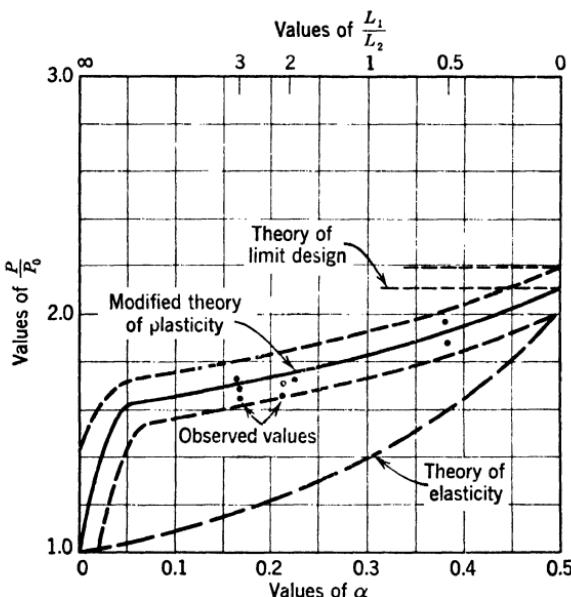


FIG. 80·2 Comparison of the results of tests of the carrying capacity of the structure in Fig. 80·1 with the carrying capacity computed on the basis of various theories.

be applied only if all intermediate states of the structure of reduced redundancy are stable until the  $(n+1)$  plastic hinge is formed.

According to the foregoing example, design for plasticity under certain conditions may be not less inadequate than the conventional design for elasticity, only less safe. However, even such rather complex conditions are fairly well met by the simple assumption of a mid-point value of carrying capacity between the extremes.

As pointed out in Art. 77, the discrepancy between the results of the theory of plastic redistribution of bending moments—

based on the assumption of the gradual spreading of isotropic plastic regions and leading to a full equalization of the bending moments carried by the critical sections of the structure—and the results of many tests<sup>80·1</sup> is due to the fact that, particularly in mild-steel sections subject to bending moments, the initiation of plastic deformation may be so long delayed that the process of transition from elastic to plastic deformation is no longer gradual but sharply discontinuous. Since the sudden plastic deformation proceeds in this case in an anisotropic manner along well-defined and rather widely spaced glide planes the development of which has been delayed by the inhomogeneity of the stress field (see Art. 75), the assumption of gradual redistribution of moments leading to full equalization of the resistance to the applied load of the critical sections can no longer be valid. The limiting carrying capacity of a mild-steel structure computed on the basis of full equalization of the bending moments in the critical sections, in general, therefore, cannot be attained; the difference between the really attainable and the limiting carrying capacity is a function of the type of the structure and the loading conditions; the more nearly equal the moments in the critical sections in the elastic stage, the more perfect the moment equalization that can be achieved by plastic redistribution.<sup>80·2</sup> Since evidently no moment equalization is required when the elastic moments are initially equal, the foregoing conclusion simply expresses the fact that the moment equalization by plastic deformation will be the more perfect, the less the plastic redistribution required to achieve it. This fact is clearly illustrated by Fig. 80·2.

### 81. Stability in Compression within the Elastic–Plastic Range

The design of axially compressed members for stability within the elastic–plastic range is usually based on the assumption of a linear work-hardening relation from which either the tangent-modulus or Engesser equation,

$$P'_{\text{B}} = \frac{\pi^2 H' I}{l^2} \quad (81 \cdot 1)$$

or the double-modulus or Kármán equation,

$$P''_{\text{B}} = \frac{\pi^2 T I}{l^2} \quad (81 \cdot 2)$$

can be derived;  $P_B$  denotes the limit buckling load,  $I$  the smallest moment of inertia of the cross section,  $l$  the buckling length,  $H' = 3H = ds/de$  the tangent modulus of work hardening in axial compression, and  $T$  the combined or Kármán modulus.

According to Engesser's generalization<sup>81·1</sup> of the Euler formula,  $P_B = \pi^2 EI/l^2$  for the elastic range, the smallest axial load at which instability of equilibrium is reached, because of the existence of an equilibrium position infinitesimally near to the straight equilibrium position, is given by Euler's equation in which the elastic modulus has been replaced by the tangent modulus or,

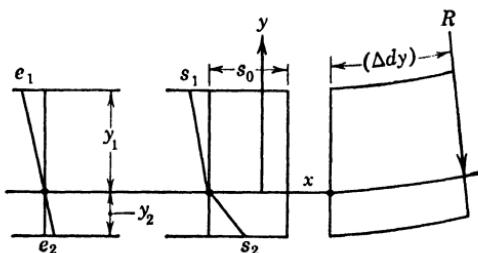


FIG. 81·1 Stress and strain distribution under maximum buckling load within the elastic-plastic range.

in case of a linear work-hardening law, by the work-hardening modulus  $H'$ , as indicated by eq. 81·1. The Engesser load is, however, not the theoretical maximum load which the column is able to support without large deflections; the ultimate load has been derived by v. Kármán<sup>81·2</sup> under the assumption of a linear work-hardening relation.

If it is assumed that, under the action of the maximum stable axial force, slight bending starts, the distribution of stresses over the section can be represented by Fig. 81·1; the uniform compressive stress due to the axial load  $s_0$  is increased at the inner edge of the bent bar by  $s_1$  and reduced at the outer edge by  $s_2$ , where the stress increase  $s_1$  is related to the deformation  $e_1$  by the relation,

$$s_1 = s_0 + 3He_1 = s_0 + H'e_1 \quad (81\cdot3)$$

while the stress reduction (unloading stress)  $s_2$  is evidently governed by the elastic modulus and related to the strain  $e_2$  by

$$s_2 = s_0 - 3Ge_2 - Ee_2 \quad (81\cdot4)$$

Under the assumption that plane sections remain plane after deformation, the following relations are valid:

$$\Delta d\phi = \frac{1}{R} = \frac{e_1}{y_1} = \frac{(s_1 - s_0)}{H'y_1} = \frac{e_2}{y_2} = \frac{(s_2 - s_0)}{Ey_2} \quad (81 \cdot 5)$$

Introduction of eq. 81·5 into the equilibrium condition,

$$\int_A (s - s_0) dA = 0 \quad (81 \cdot 6)$$

leads to the relation,

$$\begin{aligned} \int_{A_1} (s_1 - s_0) dA + \int_{A_2} (s_2 - s_0) dA &= \frac{s_1 - s_0}{y_1} \int_{A_1} y dA \\ &+ \frac{s_2 - s_0}{y_2} \int_{A_2} y dA = \frac{1}{R} (H'S_1 + ES_2) \end{aligned} \quad (81 \cdot 7)$$

where  $S_1$  and  $S_2$  denote the statical moments of the areas  $A_1$  and  $A_2$  about the neutral axis, respectively.

The bending moment,

$$\begin{aligned} M &= \int_{A_1} (s_1 - s_0)y dA + \int_{A_2} (s_2 - s_0)y dA = \\ &\frac{s_1 - s_0}{y_1} I_1 + \frac{s_2 - s_0}{y_2} I_2 = \frac{H'}{R} I_1 + \frac{E}{R} I_2 = \frac{1}{R} TI \end{aligned} \quad (81 \cdot 8)$$

where  $T$  is the Kármán modulus,

$$T = H' \frac{I_1}{I} + E \frac{I_2}{I} \quad (81 \cdot 9)$$

$I_1$ ,  $I_2$ , and  $I$  denote, respectively, the moments of inertia of the areas  $A_1$ ,  $A_2$ , and  $A = A_1 + A_2$  about the neutral axis.

The differential eq. 81·8,

$$M = \frac{1}{R} IT = -Py = \frac{d^2y}{dx^2} IT \quad (81 \cdot 10)$$

is a form identical with the Euler equation, and its solution leads to the buckling load given by eq. 81·2.

Shanley has recently shown by experiment<sup>81·3</sup> that if the tangent modulus remains constant both the Engesser and the Kármán "buckling loads" represent limiting values, the Engesser load being the lowest and the Kármán load the highest load at

which bifurcation of equilibrium positions can occur, that is, at which the column can assume a bent as well as a straight equilibrium form. For loads between the Engesser and the Kármán load the transition from the straight to the bent position requires an increase of the axial load; at the Kármán load this transition occurs at constant load. Since the actual increase of the buckling load beyond the Engesser load is relatively small and is accompanied by appreciable deflections, it may be expedient to use the Engesser tangent modulus eq. 81·1 in actual design for stability beyond the elastic range, although it should be realized that the Engesser load does not represent a genuine buckling load (limit of instability) but only a bifurcation point of equilibrium. Because the load can still increase while the column bends, it is possible to avoid elastic unloading of the section, so that the basic assumption of the double-modulus theory is no longer justified.

The difference between the Engesser and the Kármán load is very pronounced in the "reversal of strain" method of buckling observation in which the strains on opposite faces at midheight of the loaded column are plotted against the applied load. Before the Engesser load is attained the strains increase on both faces, at first at the same rate until the column starts to bend laterally, later at different rates. At the Engesser load the strain at the face with concave curvature decreases, while the other strain increases (Fig. 81·2). The maximum (Kármán) load at which the strain increases rapidly at constant load is higher than the point of strain reversal defining the Engesser load.

The application of the tangent-modulus theory is simple only if it can be assumed that the behavior of the material in uniaxial compression can be represented by a linear work-hardening relation, that is, that the tangent modulus remains constant. For a gradually decreasing tangent modulus the value of  $H'$  in

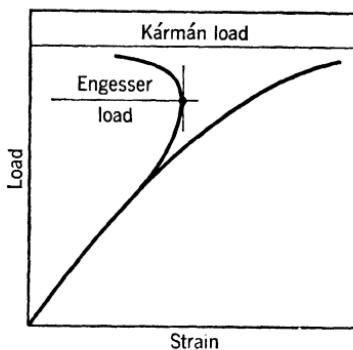


FIG. 81·2 Load-strain relations for opposite faces of column in buckling test.

eq. 81·1 depends on  $P'_B$  so that the Engesser load can not easily be determined.

The solution of the problem of stability beyond the elastic range of thin plates is considerably more difficult than that of the buckling bar because of the two-dimensional state of stress. The effect of the state of stress has two aspects which are independent of each other although they are frequently confused:

1. In plastic buckling of plates a state of stress due to the bending of the plate from the straight toward the bent equilibrium position is superimposed on the initial state of homogeneous compression. Hence the ratio of the deviator components does not remain constant, as required by the condition of isotropic plastic deformation (eq. 42·8) but changes during the buckling process as the bending stresses increase. The plastic deformation is therefore no longer isotropic but depends on the entire strain history. Bijlaard,<sup>81·4</sup> who has investigated this problem very extensively, has shown that the resistance to buckling is increased by this change of the ratio between the deviator components and that this excess of the real buckling load over the buckling load, computed on the basis of the assumption of isotropic plasticity by introducing essentially a reduced modulus of elasticity into the elastic solution,<sup>81·5</sup> is the more pronounced, the more the final ratio of the deviator components differs from the initial ratio.

2. In the state of homogeneous compression leading to buckling, the initiation of plastic deformation by the development of anisotropic flow along one or a number of visible glide planes may affect the buckling load more significantly than the usually assumed isotropic plastic deformation that could only develop as the plastic strain increases further. In this later stage, however, the isotropy of the deformation is considerably modified by the effect of the superimposed bending as outlined under 1. Hence, the evaluation of the buckling load requires the introduction of the assumption of an inherently anisotropic material in which the initiation of slip along glide planes depends not only on a stress invariant, as in the isotropic material, but also on orientation of the considered body with regard to the directions of principal shear of the applied stress field. Under this condition a pure rotation of the stress field does not represent a *neutral* change of state, as in the isotropic plastic body (see Art. 41),

but may be associated with plastic deformation along potential glide planes that have become operative as a result of the rotation. The theory of anisotropic plasticity, which has been proposed by Boeker<sup>81-6</sup> and by Brandtzaeg<sup>81-7</sup> has recently been revived and applied to buckling of plates.<sup>81-8</sup> The results of this theory, however, can be significant only when anisotropic plastic deformation by gliding along visible glide planes represents the characteristic behavior of the material at the transition between elastic and inelastic deformation.

### References

- 77·1 A. M. FREUDENTHAL, *Trans. ASCE* **112** (1947) 125.
- 78·1 F. BLEICH, *Prelim. Rept. 2d Congr. Intern. Assoc. Bridge & Structural Eng.*, W. Ernst & Son, Berlin (1936) 137.
- 78·2 W. PRAGER, *J. Applied Mech.* **17** (1950) A.
- 78·3 E. CHWALLA, *Pub. Intern. Assoc. Bridge & Structural Eng.* **2** (1933-34) 97.
- 78·4 M. GRUENING, *Tragfaehigkeit statisch unbestimmter Tragwerke aus Stahl etc.*, J. Springer, Berlin (1926).
- 78·5 J. FRITSCHE, *Z. angew Math. & Mech.* **11** (1931) 176.
- 79·1 W. PRAGER, *Bauingenieur* **14** (1933) 65.
- 80·1 F. STUESSI and C. F. KOLLBRUNNER, *Bautechnik* **13** (1935) 264.  
H. MAIER-LEIBNIZ, *Prelim. Rept. 2d Congr. Intern. Assoc. Bridge & Structural Eng.*, W. Ernst & Son, Berlin (1936) 103; *Final Rept.* (1938) 70.
- 80·2 J. FRITSCHE, *Pub. Intern. Assoc. Bridge & Structural Eng.* **6** (1940-41) 102.
- 81·1 F. ENGESELLER, *Schweiz. Bauz.* **26** (1895) 24; *Z. Ver. deut. Ing.* **42** (1898) 927.
- 81·2 TH.V. KÁRMÁN, *Physik Z.* **9** (1908) 138; *Mitt. Forschungsarb. Ingenieurwesen* **81** (1910).
- 81·3 F. R. SHANLEY, *J. Aeronaut. Sci.* **14** (1947) 261.
- 81·4 P. P. BIJLAARD, *Pub. Intern. Assoc. Bridge & Structural Eng.* **6** (1940-41) 45; **8** (1947) 17.
- 81·5 E. CHWALLA, *Ingenieurarchiv* **5** (1934) 54.
- 81·6 R. BOEKER, *Forschungsarb. Ingenieurwesen* **175-176** (1915).
- 81·7 A. BRANDTZAEG, *Failure of a Material Composed of Non-Isotropic Elements*, Skrifter Norske Videnskabers Telskaba, Trondhjem **2** (1927).
- 81·8 P. P. BIJLAARD, *op. cit.*, **6** (1940-41) 27.  
S. B. BATDORF, *Inst. Aeronaut. Sci. Preprint* 200 (1949).

## C H A P T E R

# 17

### DESIGN FOR CREEP

#### 82. Metallic Structures and Parts at High Temperatures

Design for creep of metallic parts is essentially a design for high-temperature service, since, with the exception of lead and lead alloys, engineering metals do not creep appreciably at and below room temperature. In design for plasticity the inelastic deformations are specified to remain of the order of magnitude of the elastic deformations. However, at elevated temperatures creep during the normal period of service may produce deformations of a different order of magnitude, and the designer must make the necessary provisions to allow for irrecoverable deformation of at least this magnitude, without danger of causing *functional damage*, that is, without interference with the satisfactory operation of the machinery or the structure. This deformational aspect is, however, only one aspect of the design for creep. The other aspect is that of structural damage due to the deterioration of the cohesive strength of the material as a function of the period of load application and of the temperature history (see Art. 49).

For power machinery operating at elevated temperatures up to about 850°F and for metals developed for such service temperatures, which usually have creep rates of an order of magnitude not exceeding  $10^{-7}$  to  $10^{-8}$  per hr under conventional service stresses, design for creep is mainly concerned with the deformational aspect, that is, with "functional damage" only; the specified limiting strain is usually of the order of magnitude of 0.1 percent and thus presupposes service periods of the order

of magnitude of  $10^4$  to  $10^5$  hr. In this case the limiting condition for creep design is specified in terms of functional damage; the creep aspect is irrelevant with regard to structural damage and, therefore, with regard to the specification of the limiting condition for such damage, so that design for structural safety means design for elasticity or for plasticity.

The recent development of equipment for operating service at temperatures considerably above  $850^{\circ}\text{F}$  such as gas turbines and jet engines, however, has completely changed the relative importance of the design for creep and that for plasticity or time-independent fracture and has made design for creep the principal design aspect with regard to both functional and structural damage. At temperatures above  $1000^{\circ}\text{F}$ , creep during the period of service can no longer be restricted to the order of magnitude of 0.1 percent; the design must be sound, both functionally and structurally, for total creep of the order of magnitude of about 0.5 to 1.0 percent which, with the usually considered service periods of a few thousand hours, permits creep rates of up to  $10^{-5}$  per hr. It is, moreover, no longer the functional problem of deformation limits and clearances alone that has to be considered; the reduction of fracture stress with time during the expected period of service emerges as the primary factor in creep design, whereas the importance of the effect of time-independent plasticity is very much reduced. Hence, the use of metals at service temperatures exceeding  $1000^{\circ}\text{F}$  necessitates a basic change of the design concepts; design for finite life and for appreciable permanent deformation replaces the design for time-independent carrying capacity and small deformation. Thus, stress alone is no longer the major consideration that it is in conventional design; the relevant characteristics in high-temperature design are *stress, permanent deformation, temperature, and time at a certain temperature*. It is no longer sufficient to compare the design stress with the strength of the material; but the *stress-temperature-time* history of the structure or the part must be compared with the *strength-temperature-time* history and the *deformation-temperature-time* history of the material under conditions similar to those in service.

Because of the effect at elevated temperatures of temperature and time on both deformation and strength, the operating temperature and life of the designed part must be carefully specified

in advance if the most suitable material is to be selected for the particular purpose. Since the creep-fracture stress of various metals at various temperatures decreases with time at different rates, an incorrect specification of the operating period and temperature may easily result in the selection of a material for the particular conditions that is inferior to alternative materials. Figure 56·1 shows the creep-fracture stress for certain alloys developed for high-temperature service as function of temperature with a specified time to fracture, and as function of time at a specified temperature. It can be clearly seen that the alloy that is decidedly superior above a certain temperature or below a certain operating time may become inferior for service at a lower temperature or for longer operating times if it loses its strength more rapidly with increasing temperature and time than an alternative material. This consideration is of importance not only in the selection of the most adequate alloy from existing test data, but also in the interpretation of the results of full-scale development tests of the apparatus. Since such tests may be of shorter duration than the specific operating period, and temperatures may be less controlled, the performance in this test of a particular alloy may lead to wholly unjustified conclusions concerning its performance in service.

The relative importance of plasticity in design for service conditions under which creep is the dominating factor depends on the actual character of the creep. If creep is the expression of essentially viscous intercrystalline deformation and crystal rotation along the crystal boundaries without or with very restricted slip and crystal fragmentation, the effects of plasticity are negligible since the deformation is essentially of a viscous type. In this case the total deformation, due to the deformation of the intercrystalline material, is relatively small, and fracture appears brittle. On the other hand, if creep is associated with considerable slip and crystal fragmentation, plasticity effects are of considerable importance. Under such conditions deformation prior to fracture is relatively large, and fracture appears ductile.

The necessity of differentiating between predominantly viscous and predominantly plastic creep in design arises from the different influence of either type of creep on the stress distribution and on the condition of fracture. The behavior of metals within a range of predominantly viscous creep will necessarily be very

similar to the behavior of viscoelastic materials. Hence, as a result of the elastic-viscoelastic analogy (see Art. 35) the stress distribution will not differ substantially from that of an elastic material. Substantial relief of stress concentrations, therefore, must not be expected or, at best, will be relatively small unless the nonlinearity of the creep is very pronounced; methods of design for plasticity are therefore not applicable. The resistance to structural damage should be evaluated on the basis of the elastic stress distribution, without appreciable relief of stress concentrations. On the other hand, when creep is largely due to slip and crystal fragmentation, the distribution of stresses and relief of stress concentrations will approach conditions existing within a plastic material; the work-hardening effects are practically eliminated by the relatively high rate of recrystallization at the elevated temperatures at which such conditions exist.

A closer approximation to the real stresses and strains in a nonlinear viscoelastic material can be obtained from the solution of the given problem under the assumption of a nonlinear steady-state stress-creep-rate relation, for instance in the form of a power law (see Chapter 15). However, such solutions can be obtained in closed form only for relatively simple problems, such as torsion, bending, or problems of rotational symmetry. The solution of more complex problems leading to nonlinear differential equations will usually be possible only by numerical methods.

The stresses and particularly the strains computed under the assumption of steady-state creep are significant only if the design is for long life, or if the range of transient creep is very short. In the design of parts for short life the analysis should consider that creep proceeds at a decreasing rate before the steady state is reached. Such analysis can proceed only by methods of numerical approximation. It can however be shown<sup>82-1</sup> that under certain conditions the strains occurring during the transient stage of creep may be much larger than those obtained by consideration of the steady-state creep rate alone.

If substantial vibratory stresses superimposed on a steady stress have to be considered, the relative importance of time effects and of cycle effect will depend on the comparative creep effects of the steady and the alternating stress. When the creep under the steady stress is essentially viscous, the cycle effect

that is associated with slip and fragmentation will be appreciable and design for stress reversal necessary. When, on the other hand, creep under the steady stress is produced by extensive slip and fragmentation, the cycle effect may be very small, since the intensified fragmentation under the vibratory stress cycles will also produce intensified recrystallization as a result of which the work hardening associated with cycle effects is largely eliminated; since in this case the difference between creep strength and fatigue strength practically vanishes (see Art. 58),

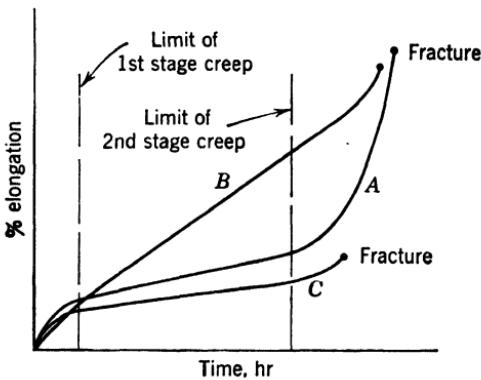


FIG. 82·1 Schematical creep curves indicating different performance (after R. B. Smith<sup>58-6</sup>).

the variable (cyclic) stress is of no significance and need not be considered in design.

A very important aspect in design for high-temperature service is the consideration of the thermal stresses arising in the designed parts as a result either of differential expansion associated with thermal gradients or of localized plastic deformation due to the relief of high compression stresses produced by local heating, and followed, on subsequent cooling, by very high residual tensile stresses. The deviatoric part of the thermal peak stresses can be relieved only by the essentially plastic type of creep, not by viscous creep. Since the highest thermal stresses are transient, as they occur at the start or at the termination of an operation of the machinery, the most important property of the material will be a very high short-time creep-fracture stress, which will prevent thermal cracking during the transient heating and cooling cycles (*thermal fatigue*).

The difference between the two types of creep in metals used for high-temperature service is clearly visible in the typical creep curves of different alloys, as shown schematically in Fig. 82-1. Although all three alloys fracture under the same stress at nearly the same time, diagram *B* is characteristic for a material the creep of which is largely due to rotation along grain boundaries, crystal slip, and fragmentation, from the very beginning of the load application to final rupture, which is preceded and accompanied by considerable elongation. Diagram *C* expresses the behavior of a metal with moderate creep, resulting mainly from intercrystalline deformation during all three stages of the creep process; fracture accompanied by small deformation is sudden and has a brittle appearance. Neither material *B* nor *C* is really satisfactory from the designer's point of view. The creep of type *B*, although accompanied by large-scale plastic relief of stress concentrations, actually represents a process of progressive damage to the cohesive strength of the metal, a process which starts at the moment of load application; the slight upward inflection of the diagram near fracture marks but a final acceleration of this process of separation, probably due to the propagation of visible cracks. There is neither a first nor a second stage of creep since the third stage extends practically over the entire time of loading. The predominantly viscous creep of type *C* is restricted in magnitude and represents true flow with no immediate real damage but an accumulation of potential damage due to grain boundary relaxation (see Art. 58); it is this gradual accumulation of potential damage that subsequently produces sudden fracture with small additional deformation after a short third stage. Plastic relief of stress concentrations under sustained loads must not be expected. Hence, although the deformational performance of the material *C* is satisfactory, its capacity to redistribute concentrated stress is not; its structural performance is therefore rather inadequate.

Although a metal whose creep behavior is represented by diagram *A* deforms essentially in a viscous manner like metal *C* during the first two stages, the character of creep changes completely during the third stage; it becomes plastic. Thus, the accumulated damage within the grain boundaries at the end of the second stage of creep is of no consequence, since fracture is preceded by considerable plastic deformation and redistribution

of stresses. The ultimate carrying capacity of the structure and its parts, therefore, can be derived on the assumption of fully plastic behavior, so that material A represents comparatively the most satisfactory type of behavior for high-temperature service.

In specifying working stresses for creep design from observed creep diagrams, the expected service life must be considered, and the stresses set up so as to ensure adequate fracture strength and a limiting deformation that would not exceed an amount

specified with regard to successful operation of the designed machinery during its period of service. Working stresses for short-life equipment such as jet engines are based on an expected life of 1000 to 2000 hr. Since creep tests of such duration are usually available, there is no difficulty in specifying the stress that would produce a creep rate and thus a final

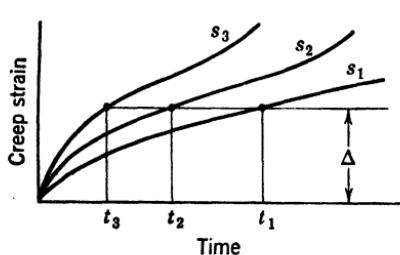


FIG. 82·2 Relation between design stress and expected life  $t$  in design for limiting creep strain  $\Delta$ . ( $s_3 > s_2 > s_1$ .)

deformation not larger than the limiting amount; for the same limiting deformation a shorter expected service life permits design with a higher working stress (Fig. 82·2). When, on the other hand, equipment such as power machinery is designed on the basis of an expected life of 100,000 hr, whereas in creep tests testing times of 10,000 hr are only very rarely exceeded, the necessary extrapolation from the 10,000 hr test is safe only if, at the selected stress level, the third stage of creep starts beyond the expected service life. Otherwise the extrapolation would result in excessive deformation occurring during the service period, associated either with progressive damage to the cohesion of the material or with actual fracture (Fig. 82·3). With regard to the use of results of creep tests for the design of machine parts, the effect of the state of stress should necessarily be considered. Since only the deviatoric part of the stresses produces creep, it cannot be expected that the creep rates in uniaxial tension as obtained from a conventional test will be comparable to the creep rates under the three-dimensional state of stress existing in most

machine parts. This consideration applies both to the creep rate and to the time at which the third stage is initiated.

If the working stresses for creep design are not governed by the limiting deformation, but by the fracture strength, a different design approach is necessary for metals with essentially viscous and metals with essentially plastic creep because of the difference in the effect of stress concentrations in either case. The fracture stress for 100,000 hr or more at full temperature must be determined by extrapolation from short-time tests which is possible

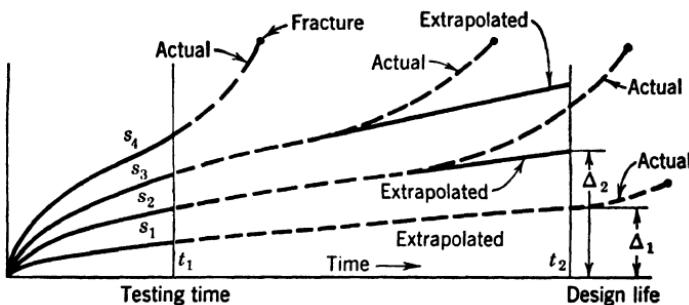


FIG. 82-3 Comparison of actually expected creep curves extrapolated from short-time creep tests for design for a limiting creep strain  $\Delta$  at a selected design stress  $s$  (extrapolation permissible only at stress  $s_1$ ).

only if the slope of the function of fracture stress versus the logarithm of life does not change abruptly.

Present practical high-temperature design is usually based on the maximum-stress theory of fracture, if general conditions of stress are encountered. There appears to be no factual information available on the basis of which a rational condition of fracture of metals at high temperatures under a general state of stress could be derived. Where creep is restricted and constitutes a secondary, mainly deformational effect, as in design for service temperatures not exceeding 850°F, the fracture criterion developed in Art. 59 remains more or less valid. However, where creep and creep-rupture become the main problem, as in design for service temperatures above 850°F, part of the inelastic deformation is the expression of progressive damage instead of being only an indication of a change of structure by slip and fragmentation. Hence, the condition of fracture can no longer be expressed in terms of elastic strain energy alone; the limiting amount of

this energy probably becomes a function of the creep preceding fracture.

Whether creep is essentially viscous or essentially plastic depends mainly on stress level and temperature. An increase of stress or a change of temperature may therefore change the character of creep as well as of the creep fracture. Thus, correlation between design stress and ultimate or fracture stress is rather meaningless, as either stress refers to a different duration of the load and thus, in certain cases, may refer to a different type of deformation and fracture. By selecting the design stress in terms of a certain percentage of the fracture stress for a certain duration of the load, a considerable safety margin is actually introduced with regard to the period of operation, since fracture at the design stress is not made impossible or less probable than at the ultimate stress, but only requires a longer period of application of the load. Because of this extension of the time to fracture, the fracture phenomenon itself may be changed from ductile, owing to essentially plastic creep, to brittle owing to essentially viscous creep, or vice versa, depending on the character of the material and on the temperature. Thus the approach to the design at both stress levels is no longer identical. These facts make any rational interpretation of the concept of working stress in design for high-temperature creep almost impossible. It appears therefore reasonable to specify that such design should be based on judicious joint evaluation of the expected most critical loading and temperature and the maximum time of operation rather than on service conditions, comparing the maximum stresses and times directly with the observed fracture-stress-time diagram of the material. However, even this procedure is difficult, unless only a single operating temperature is considered; no information is at present available concerning the damaging effect of a time sequence of temperature cycles, such as, for instance, overheating.

The concept of the safety factor in design for high-temperature creep strength is considerably more complex than the same concept in design for elasticity or plasticity. This complexity results from the interrelation among time, temperature, and strength which introduces the duration of the load at certain temperatures as a significant characteristic of this load. Since the effect of load fluctuations is to be combined with temperature

fluctuations of various duration, it cannot be interpreted in terms of the required safety without there being considered simultaneously the expected durations of the various load intensities at the various temperatures and their cumulative effect on the fracture strength of the material. Fluctuations of load intensity are therefore not independent of the fluctuations of resistance so that the simple statistical concept proposed in Art. 77 is not applicable.

### 83. Viscoelastic Materials

In the design of parts and structures of viscoelastic materials the fact that the amount of deformation or of strain is not a direct indication of the intensity of stress is of principal importance. In such materials a nonlinear force-deformation diagram obtained in a tension test is not an indication of a nonlinear stress distribution in bending, since nonlinear behavior may be a time effect only. In Art. 35 it has been shown that a linear viscoelastic body in tension or shear is characterized by stress-strain or force-deformation diagrams of a curvature which depends on the applied strain rate. The stress distribution in bending of a beam of such material, however, is linear as long as the relation between stress and *rate* of strain remains linear. Thus, nonlinearity of a stress-strain diagram in a mechanical test is significant for design only if the nonlinear behavior is not produced by time effects alone and does, therefore, not change appreciably with the applied rate of strain.

Design for strength of linear or nearly linear viscoelastic materials should therefore be based on the stress distribution given by the theory of elasticity. No matter what the inelastic deformation observed after the forces have been applied, the distribution of stresses in structures of such (volume-constant) materials does not change but remains essentially identical with the elastic distribution produced immediately on load application. Observed inelastic "yielding" is therefore not an indication of relief of stress concentrations, if "yielding" is used as a synonym for creep.

Considering creep in the design of redundant structures, the difference in the effect of creep on stresses produced by loads (*load stresses*) and on stresses produced by an initial deformation (*deformation stresses*) necessarily influences the design procedure.

In redundant structures of linear or nearly linear viscoelastic behavior and within the range of small deformations the distribution of load stresses and the values of the redundants are proportional to the applied loads and independent of deformation; the results of elastic analysis of such structures are therefore valid, in spite of their inelastic behavior. No redistribution of values of redundants takes place as a result of irrecoverable deformation. However, if the stresses or redundants result from an initially applied deformation, they will tend to fade out asymptotically because of the relaxation effects in the material. This is immediately evident from the linear viscoelastic equations of bending considered as a uniaxial problem.

The equation of an element of a Maxwell body for uniaxial stress is (see eq. 35·29)

$$\dot{\epsilon} = \frac{1}{E} s + \frac{1}{\lambda} \dot{s} \quad (83 \cdot 1)$$

The approximate relation between the bending moment  $M$  and the curvature  $1/R \sim -d^2w/dx^2$  which for the elastic material is

$$\frac{d^2w}{dx^2} = - \frac{1}{EI} M \quad (83 \cdot 2)$$

becomes, on differentiation with regard to time,

$$\frac{\partial}{\partial t} \left( \frac{\partial^2 w}{\partial x^2} \right) = - \frac{1}{EI} \cdot \frac{\partial M}{\partial t} \quad (83 \cdot 3)$$

For the ideal viscous beam, this relation is obtained directly from the viscoelastic analogy:

$$\frac{\partial}{\partial t} \left( \frac{\partial^2 w}{\partial x^2} \right) = - \frac{1}{\lambda I} M \quad (83 \cdot 4)$$

By carrying out the operation indicated by eq. 83·1. the equation is obtained,

$$\frac{\partial M}{\partial t} + \frac{M}{\tau} + EI \frac{\partial}{\partial t} \left( \frac{\partial^2 w}{\partial x^2} \right) = 0 \quad (83 \cdot 5)$$

with the relaxation time  $\tau = \lambda/E$ .

The solution of eq. 83·5 is

$$M = e^{-t/\tau} \left[ M_0 + EI \int_0^t e^{t-\tau} \frac{\partial}{\partial t} \left( \frac{\partial^2 w}{\partial x^2} \right) dt \right] \quad (83 \cdot 6)$$

For an impressed deformation the curvature remains constant, so that the right-side integral of eq. 83·6 vanishes, and the initially produced moment  $M_0$  decreases towards zero according to the function  $M_0 e^{-t/\tau}$ .

The differences and similarities in the behavior of redundant metal structures and redundant viscoelastic structures and their effect on design can be illustrated by considering the continuous beam over two spans, loaded by symmetrical concentrated loads at mid-span, which produce a settlement of the central support

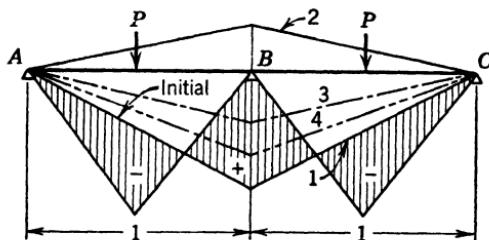


FIG. 83·1 Change of line of support moment of viscoelastic beam over two spans with yielding central support B. (1, level support; 2, instantaneously on yielding of support; 3 and 4, at different times after yielding).

by a finite amount. It has been found in Art. 79 that the carrying capacity of such a metal structure is determined by the plastic resistance of the critical sections near mid-span and over the central support and is affected neither by the value of the redundant in the elastic state nor by the (small) yielding of the support. In the case of the linear viscoelastic structure the distribution of stresses and the value of the redundant are identical with those of the elastic structure. Yielding of the support is accompanied by an immediate change in the value of the redundant and the stress distribution. These changes, however, are not permanent since they are the result of an imposed deformation but vanish with time, and the structure tends asymptotically towards the elastic distribution of redundants and stresses. The design of such structures for strength should therefore be based on their elastic analysis, design for creep remaining a secondary purely deformational aspect. These conclusions are borne out by experiments with resins<sup>83·1</sup> (Fig. 83·1).

The creep aspect becomes of primary importance where the distribution of stresses and the values of the redundants depend

on the deformation itself, since the structure can no longer be analyzed in its undeformed condition. The influence of the deformation on the stresses and the limitation of the carrying capacity by the existence of a stable deformed shape change the strength problem into a stability problem. The fact that the stability of such structures or parts is a function of their total deformation introduces the dependence of this stability on the amount of creep and thus on creep rate and the duration of the load; the limit of structural stability becomes therefore a function of time. The principal structural forms in the design of which these considerations are of practical significance are columns and arches.

**COLUMNS.** The differential equation of the viscoelastic column with an initial deviation from the straight line  $w_i$  is obtained by introducing the bending moment due to the axial compressive force  $P$ ,

$$M = P(w_i + w) \quad (83 \cdot 7)$$

into eq. 83·5; the resulting equation has the form,

$$EI \frac{\partial}{\partial t} \left( \frac{\partial^2 w}{\partial x^2} \right) + P \left( \frac{w}{t} + \frac{w}{\tau} \right) = -P \frac{w_i}{\tau} \quad (83 \cdot 8)$$

For  $t = 0$  the solution  $w = w_0$ , where  $w_0$  denotes the solution of the differential equation for elastic buckling. Equation 83·8 has an analytical solution of the form  $w = X(x)T(t)$  only if  $w/w_i = \text{const}$ , that is, if the time-dependent deflection  $w$  is proportional to the initial deviation  $w_i$ . Although in this case the theoretical ultimate buckling load  $P = P_{B0}$  is that of the elastic column, very large deflections are reached under any load  $P < P_{B0}$ , if its time of application is long enough.

If  $w$  and  $w_i$  are not congruent, eq. 83·9 can be solved by successive approximation.<sup>83·2</sup> Considering a compressed strut with lateral load producing a moment distribution  $M_{x0}$ , a first approximation  $w_i$  of the deflection is obtained by integrating eq. 83·9 with  $M = M_{x0}$ . The second and the following approximations  $w_2, \dots, w_n$  are obtained by solving eq. 83·9 successively with  $M_1 = Pw_1$ ,  $M_2 = Pw_2 \dots M_{n-1} = P \cdot w_{n-1}$ . The solution  $w = w_1 + w_2 + \dots + w_n$  is a power series in  $(P/P_{B0})$  and  $(t/\tau)$ ; it converges only if the duration of the load does not exceed a critical time  $t_0$ . For  $t \geq t_0$  the series becomes divergent; the

deformation thus tends towards an infinite value for any load  $P > 0$ , if its duration exceeds  $t = t_0$ , the *buckling time*; as long as  $t < t_0$ , the equilibrium remains stable. The limiting relation  $P/P_{B0}$  versus  $t_0/\tau$  is obtained from the condition that consecutive terms of the power series

$w(P/P_{B0})$  decrease in value.

Because of the large deformations that must be expected under loads even below the theoretical instability limit, the practical buckling loads will be considerably below the theoretical limiting loads. Figure 83·2 shows the buckling-load versus buckling-time curve for a compressed strut with constant initial eccentricity  $w_i = \text{const}$ , computed by the outlined method of successive approximation. The reduction of the stability with time is shown to be considerable and would require a different treatment in design of viscoelastic struts with

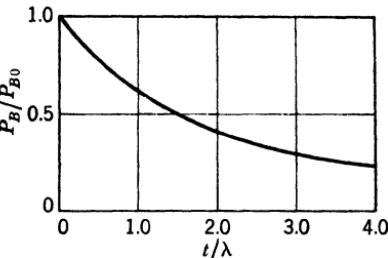


FIG. 83·2 Buckling-load versus buckling-time diagram of viscoelastic strut.

regard to transient and to sustained compressive forces.

The theoretical conclusions are borne out by experiments, as can be inferred from Fig. 83·3 in which the observed buckling strength of a strut of phenolic resin of slenderness ratio 75 is shown as a function of the buckling time.<sup>83·3</sup>

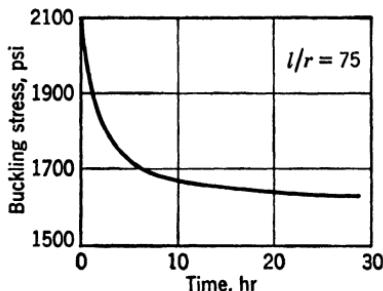


FIG. 83·3 Buckling-stress versus time curve of phenolic resin strut (after Ross<sup>83·3</sup>).

structure as long as the total deformations may be assumed to remain negligibly small with respect to the rise of the arch. In this case the thrust as well as the load stresses are unaffected by creep, whereas the deformation stresses and the differential thrust due to rib shortening or to yielding or rotation of supports

ARCHES. The behavior of redundant arches of viscoelastic material is identical with that of any other type of redundant

tend to vanish with time, asymptotically re-establishing the elastic distribution of stresses and redundants due to loads only. Thus, if an unforeseen yielding of abutments occurs, accompanied by a reduction of the initial thrust due to the load, this reduction will be compensated in the course of time, and the initial condition of the elastic arch, envisaged in the design, will be re-established. Hence, it appears unnecessary to anticipate and allow for a small movement of supports in the design of viscoelastic arches, if the material is able to resist the transient stresses, which appear at the time of the movement.

Arches that are so slender, flat, and heavily loaded that their change in configuration due to the deformation can no longer be neglected in the analysis of redundants and stresses do not behave in the manner just outlined.

If the bending moment is so much affected by the deflection  $w$  that this influence cannot be neglected, the right-hand term of eq. 83.2 becomes

$$M = M_0 - H(y - w) - M_L + \frac{x}{l}(M_L - M_R) \quad (83.9)$$

where  $M_0$  denotes the moment of the simply supported span  $l$ ,  $H$  the horizontal thrust, and  $M_L$  and  $M_R$  the fixed-end moments at the left and right abutment, respectively. Hence, the differential equation of the deflection curve,

$$\frac{d^2w}{dx^2} + c^2w = A + B\left(\frac{x}{l}\right) + \dots + N\left(\frac{x}{l}\right)^n = X\left(\frac{x}{l}\right) \quad (83.10)$$

where the right-hand side represents the development of eq. 83.9 into a power series, the constants of which depend on applied load, conditions of support, and variation of moments of inertia along the rib; the constant

$$c^2 \doteq \frac{H}{EI_c} \quad (83.11)$$

where  $I_c$  denotes the moment of inertia of the crown section. The solution of eq. 83.10 is

$$w = C_1 \cos cx + C_2 \sin cx + F(x) \quad (83.12)$$

where  $F(x)$  is a particular polynomial solution of the same order as  $X(x/l)$ , the coefficients of which are determined by introducing

$F(x)$  into eq. 83·10 and comparing it term by term with  $X(x/l)$ ; the coefficients are functions of the redundants  $H$ ,  $M_L$ , and  $M_R$ . The redundant moments are defined by the two boundary conditions,

$$\left(\frac{dw}{dx}\right)_{x=0} = \phi_1 \quad \left(\frac{dw}{dx}\right)_{x=1} = \phi_2 \quad (83\cdot13)$$

where  $\phi_1$  and  $\phi_2$  denote the imposed angular rotations of the abutment sections; the thrust may be determined from the condition that the horizontal projection of the deformed rib is equal to the span.

$$\int_0^l (ds + \Delta ds) \cos (\phi + \Delta d\phi) = l \quad (83\cdot14)$$

For the fixed-end arch with unyielding abutments  $\phi_1 = \phi_2 = 0$ ; for the two-hinged arch  $M_L = M_R = 0$ .

The bending moments in the rib are determined by the expression:

$$M = H(C_1 \cos cx + C_2 \sin cx) + EI_c \frac{d^2}{dx^2} F(x) \quad (83\cdot15)$$

Evidently the superposition principle is no longer valid, and influence lines cannot be used. In order to determine the critical values of the combination of bending moments and thrusts, eqs. 83·10 and 83·14 must be solved for the most unfavorable loading conditions.

The difference between the results of the conventional arch theory and the outlined deflection theory which is a function of the ratio of dead and live load and the rigidity of the rib expressed by the arch characteristic ( $cl$ ) is illustrated in Table 83·1, giving the values of bending moments at the quarter-point of the span of a two-hinged flat parabolic arch of a rise to span ratio 1/9, subject to dead load  $p_w$  and live load  $p$  extending over one-half the span from abutment to crown, for different ratios  $p_w/p$  and two arch characteristics ( $cl$ ) = 2 and ( $cl$ ) = 5 computed by the deflection theory.<sup>83·4</sup>

TABLE 83·1

$p_w/p$	1	6	12
$cl = 2$	$1/47.5$	$1/44$	$1/35.5 pl^2$
$cl = 5$	$1/30$	$1/26$	$1/22.5 pl^2$

The relation between the bending moments of the arch and the characteristic ( $cl$ ) indicated in Table 83·1 explains the considerable influence of creep in the design of arches of viscoelastic material. Since

$$cl = l \sqrt{\frac{H}{EI_c}} = \frac{l}{r} \sqrt{\frac{s_0}{E}} \quad (83 \cdot 16)$$

where  $r$  denotes the radius of inertia of the crown section, the bending stresses in the arch depend on the slenderness ratio  $l/r$ , on the sustained uniform compression stress  $s_0$ , and on the elastic modulus. In first approximation the effect of creep may be introduced as a gradual reduction of the apparent elastic modulus (secant modulus)  $E'$  or, according to eq. 48·1,

$$E' = E \frac{1}{1 + \frac{t}{\tau}} \quad (83 \cdot 17)$$

where  $E$  denotes the initial (elastic) modulus. Introducing eq. 83·17 into 83·16 gives

$$cl = \frac{l}{r} e_0 \sqrt{1 + \frac{t}{\tau}} \quad (83 \cdot 18)$$

where  $e_0 = \sqrt{s_0/E}$ . With ( $cl$ ) thus becoming a function of time, the bending moments of the arch will necessarily increase with time.

This increase will affect primarily the dead-load moments, since no appreciable creep and therefore no apparent reduction of the elastic modulus is caused by transient loads. The slight influence of creep on the moments due to service loads is only a result of the fact that, according to eq. 83·9, the bending moments produced by any load depend on the total thrust, including the dead-load thrust, and on the total deflection, including the deflection due to dead load, and that both dead-load thrust and deflection depend on creep.

The comparison of the behavior of arches with characteristics below  $cl = 2$  and those with characteristics  $cl > 2$  shows the considerable difference of the effects of viscous creep on either type of structure.<sup>83·5</sup> Whereas for  $cl < 2$  the stress problem is significant, the creep effect being beneficial with regard to defor-

mation stresses and negligible with regard to load stresses, it is damaging for arches with  $cl > 2$ , as the significance of the stress problem is gradually reduced and the stability problem becomes increasingly significant. Figure 83·4 shows the effect of the end conditions of the arch on its creep sensitivity.

#### 84. Concrete and Reinforced Concrete

Concrete is formed by an aggregation of loose grains (sand and aggregate) held together by a highly viscous liquid, the cement paste. The viscosity of this liquid increases with time, as a result of chemical changes within the structure (crystallization) until a complete crystalline network blocks all viscous deformation. The relative volume of grains and of viscous medium determines the mechanical behavior of the concrete. Hardened cement paste is essentially a (nonlinear) Maxwell body of high viscosity and cohesion; the aggregates and sand form a noncohesive, granular mass, the resistance of which to irrecoverable deformation by shear (in this case identical with "gliding rupture") is the result of friction between the grains, which in turn is a function of the applied hydrostatic pressure. The observed behavior of concrete is enclosed between these two limits. The high-grade concretes, which are very rich in cement, deform essentially like viscoelastic materials with a viscosity of the order of magnitude of about  $10^{16}$  poises. The larger the amount of aggregates, the less pronounced the effects of viscosity and of cohesion and the more pronounced the effects of internal friction. Thus, in rich concretes the distribution of stresses is essentially elastic, whereas in the incoherent granular mass the stress at which deformation occurs and proceeds remains constant, as in a plastic material. It may therefore be assumed that various

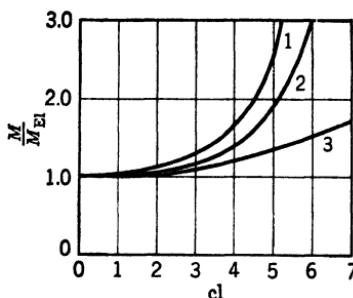


FIG. 83·4 Bending moments at  $l/4$  of flat parabolic arches with different support conditions loaded by dead load and unsymmetric live load, computed by the conventional theory  $M_{EI}$  and the deflection theory as functions of the arch characteristic  $cl$ . (1, three-hinged arch; 2, two-hinged arch; 3, fixed-end arch.)<sup>83·4</sup>

mixes of concrete, defined by various ratios of aggregates and cement, will show types of behavior varying between that of the purely viscoelastic and that of the plastic material and therefore manifest various types of nonlinearity of the stress-strain relation.

There is, however, one fundamental difference between the plasticity of a metal and the apparent plasticity of a granular material under constant stress. This difference is not so much in the type of motion but in the forces acting between the particles involved in it. Whereas in plastic slip in metals the interacting (cohesive) forces are strong enough to re-form continually the cohesive bonds during the motion, the apparent cohesion between the particles of the aggregate, resulting, in fact, only from the adhesion between the particles and the cement medium, is not re-established if it has been destroyed once in the course of the relative motion of particles under the action of external forces. Hence, the effect of plasticity of a metal and of the apparent plasticity of concrete containing a relatively large volume of aggregates is of a basically different type: metal plasticity produces relief of elastic peak stresses by nonlinear behavior due to transcrystalline slip, without damage to cohesion during this process, whereas the relief of elastic peak stresses produced by the "plasticity" of the concrete is accompanied by a certain amount of local destruction of cohesion and is therefore the expression of a process of internal damage. Therefore, the observation of considerable nonlinearity of the stress distribution of a concrete section in bending which is not solely due to the nonlinearity of the creep tacitly implies that under the applied load considerable local destruction is produced within the material; otherwise, the concrete would remain a viscoelastic material with an essentially linear or slightly nonlinear stress distribution.

The basic difference between the plasticity of metals, the "plasticity" of concrete, and its viscosity is not sufficiently realized in the design of concrete and reinforced-concrete structures and in the selection and adequate evaluation of the material. It is therefore frequently assumed that the absolute or the relative amount of irrecoverable deflection of a concrete structure has some structural significance or can be used to evaluate performance or damage, as, for instance, in the case of the conventional loading test of reinforced concrete structures. There is no justification for this assumption; it is the damaging part of the

irrecoverable deformation alone, that is, the part that produces the time-independent nonlinear distribution of stress in bending, that might provide an indication of structural damage, if it could be separated from the viscous deformation, which bears no relation to damage or structural performance, but the amount of which considerably exceeds the deformation due to structural damage.

Creep in concrete proceeds at a decreasing rate. This is due partly to the stiffening (crystallization) of the cement paste with time, partly to the gradual stoppage of the viscous flow of the cement under sustained load by the particles of the aggregate coming into contact and forming a continuous rigid skeleton. The closer the initial structure of the concrete to the formation of such a skeleton, the shorter the period of creep and the smaller its amount. The duration of creep is therefore the shorter, the leaner the mix and the more elongated the stone particles.

Creep in concrete can be effectively reduced or stopped within a relatively short time by the introduction of reinforcement in the direction of the motion, since the viscous deformation of the concrete is thus coupled in parallel with the elastic or, above the yield limit, with the plastic or work-hardening deformation of the steel. The higher the yield stress of the reinforcing steel, the more effective its performance in reducing the creep of the concrete.

The design of "prestressed" reinforced-concrete structures<sup>84-1</sup> is to a considerable extent the problem of the interaction of a viscoelastic beam with an elastically restrained reinforcement which produces an initial bending moment, counteracting the moment produced by dead load and service load. Because of the high compressive stresses involved, the concrete used must be of a very high grade and therefore of very nearly viscoelastic behavior.

The "prestress" in the reinforcement, producing the counteracting moment, is introduced by imposing a constant initial deformation (shortening of the fiber) by means of straining the reinforcement before the concrete is poured and releasing the straining force after sufficient bond has developed in the hardened concrete or, if the bars are free to move in channels of the beam and are anchored by end plates, by straining them after the concrete has hardened. After hardening of the concrete the "pre-

stress" is subject to relaxation; it will therefore gradually decrease with increasing viscous deformation of the concrete. It is only because of the fact that the creep of concrete tends with a decreasing rate toward a finite limit that a considerable part of the "prestress" can be permanently retained in the structure.

### References

- 82·1 L. F. COFFIN, P. R. SHEPLER and G. S. CHERNIAK, *J. Applied Mech.* **16** (1949) 229.
- 83·1 A. D. ROSS, *Structural Engr.* **24** (1946) 421.
- 83·2 A. M. FREUDENTHAL, *Rept. 6th Intern. Congr. Applied Mech. Paris* (1946).
- 83·3 A. D. ROSS, *op. cit.*, 419.
- 83·4 A. M. FREUDENTHAL, *Pub. Intern. Assoc. Bridge & Structural Eng.* **3** (1935) 100; **4** (1936) 249.
- 83·5 *Ibid.*, **4** (1936) 249.
- 84·1 G. MAGNEL, *Prestressed Concrete*, Concrete Pub., London (1948).

## C H A P T E R

# 18

### SIGNIFICANCE OF MECHANICAL TESTING. INTERPRETATION OF RESULTS

#### **85. Mechanical Tests and Inelastic Behavior. Machine Effects**

The significance of the variables observed in mechanical tests depends on the purpose of the test. This purpose may be:

1. The determination of properties and characteristics of basic physical significance.
2. The determination of properties and characteristics significant in manufacturing processes.
3. The derivation of data for design of engineering structures.
4. The control of the uniformity of a manufacturing process in terms of the uniformity of a certain mechanical characteristic of the product.

The variables observed express either the kinematical response of the tested specimen to imposed dynamical conditions, or the dynamical response to imposed kinematical conditions, or, finally, the energetic response to the imposed, usually dynamical, testing conditions. The phenomenon observed is thus either the deformation of the specimen under a given load or the resistance to deformation of the specimen, expressed in terms of the force developed under the imposed conditions, and interpreted in terms of the stress with which the volume element opposes the external conditions. The significance of this resistance is, however, not independent of the deformation associated with it, unless the material is perfectly elastic. It is thus the interpretation of the relation between resistance and inelastic deformation

that provides the basis for an interpretation of the significance of mechanical tests.

The interpretation of the significance of the characteristics determined in a mechanical test is relatively simple, if the purpose of the test is either to determine the material characteristics that are directly significant in a manufacturing process—such as the work-hardening capacity of a metal for the deep-drawing process or the viscosity of a resin for an extrusion process—or to control the uniformity of such a process by controlling a certain specified characteristic. In the latter case the functional interrelation between the observed characteristic and any real physical property which would be relevant in the use of the material may be vague or nonexistent. It is only the *uniformity of the manufacturing process* that the tests are designed to control; therefore the interpretation of the *variation* of the observed characteristic, not of the characteristic itself, is of real significance.

The interpretation of tests, the purpose of which is the observation of basic physical properties, is, in general, rather difficult, since it is almost impossible to devise a testing procedure in which the observed quantities are basic physical properties of a material, and in which, moreover, these properties are not changed by the test itself. There is a very long way between the performance of a mechanical test and the interpretation of the observed characteristics. Thus, the observed "basic" properties of the material are usually rather complex characteristics of both the testing procedure and the material, not as it is, but as it is being changed by the test. These characteristics are therefore related to the material in its newly changed conditions, not in its initial condition. Hence, correlation between test results and the performance of the material in service, or between different types of tests, can be expected only if the changes that the material has undergone during the test are similar to the changes that it is expected to undergo in service or in the different types of test. Since mechanical changes of state can usually be interpreted or expressed in terms of inelastic deformation, the criterion of correlation between test and performance or between different tests can be formulated in terms of the similarity or of correlation of inelastic deformation under the compared conditions. Thus, unless the inelastic deformations produced in the compared tests can be correlated or are of a similar

character and order of magnitude, correlation between test results can not be expected.

This fact makes the *derivation* of design data from test results particularly difficult, since such data can only be obtained from tests that reproduce service conditions rather closely, whereas conventional tests are designed for expediency of testing rather than for reproduction of conditions of performance. Hence, conventional test results must be *interpreted* in terms of performance, a procedure the success of which again depends on the degree of similarity of inelastic behavior in service and in the test.

Similarity of inelastic behavior in various types of tests or in test and performance requires a sufficient similarity of the characteristics that affect the inelastic behavior, such as state of stress, strain rate, temperature, and previous history as expressed by inelastic strain. Since, under conditions for which an equation of state may be assumed to exist, strain rate and temperature are interchangeable, similarity of conditions resulting in similar inelastic behavior requires similarity of a combined strain-rate-temperature parameter rather than similarity of either strain rate or temperature considered separately.

In mechanical tests specimens of a material are loaded by constant or gradually increasing forces until separation occurs or considerable deformations are produced. From the recorded force-deformation or deformation-time or force-time diagrams, the characteristic properties of the materials are derived. The specimens are loaded either by applying a mass directly or with the aid of levers, changing the force by changing either the mass or the lever arm, or by moving the mass with a certain acceleration; or the specimen, set into the mechanical system of the testing apparatus, is forced to deform by specified amounts, and its resistance to the imposed deformation is measured. In interpreting the recorded relation between force and deformation, it is important to recognize the influence of the testing apparatus itself on the shape of this relation. Since the specimen is not independent but forms part of the mechanical system, the mechanical response of the entire system to either deformation or load must be considered and analyzed, if the force-deformation diagram is to be interpreted in terms of the response of the specimen itself. The characteristic of the testing apparatus, relating

the applied forces to the deformation of the apparatus or vice versa, must be known before the results of the mechanical test of a specimen obtained in this apparatus can be analyzed; this response must be considered both under static conditions, that is, successive states of equilibrium, and under the transient dynamic conditions of actual loading.

In testing machines in which the loads are applied by connecting a mass to the specimen either directly or by the aid of levers (*force machines*), inelastic deformation proceeds under constant load only if the conditions are very nearly static; that is, the deformation increases very slowly. Whenever the inelastic deformation proceeds rapidly, the inertia forces of the applied masses must be considered. The influence of the inertia forces increases with increasing lever ratio  $z$  since the moment of inertia  $I$  of the applied mass  $m$  increases as the square of this ratio, whereas the increase of the force with the ratio is only linear.

$$\text{The natural frequency of the system } \omega = \sqrt{\frac{\text{const}}{I}} = \frac{1}{z} \sqrt{\frac{\text{const}}{m}}$$

thus decreases with increasing ratio  $z$ ; since a high natural frequency determines the sensitivity of the apparatus to indicate rapid changes of deformation of the specimen, testing machines with large lever ratios are inadequate for recording rapid inelastic deformation. Hence, no information is obtainable from the load-deformation diagram observed in such machines concerning the actual process of deformation before practically static conditions have re-established themselves. However, at the conclusion of the transient (dynamic) process, the force defining the state of equilibrium has not been changed.

In testing machines in which a specified deformation is imposed mechanically or hydraulically (*deformation machines*), the resistance to deformation of specimen and apparatus including the resistance-measuring device determines the recorded force. The whole system can be considered a rather complex combination of springs, represented by the specimen, the compressible fluid, and the elasticity of the machine parts and of the load indicator, and of masses. The character of the machine is determined by the relative magnitude of the combined elastic deformation of machine-parts and liquid and the elastic deformation of the specimen. If the elastic deformation of the machine

is much larger than the deformation of the specimen, the machine appears *soft*; inelastic deformation proceeds without considerable reduction of the force, as can be inferred from Fig. 85·1 in which the specimen and the machine are represented by two springs. If, under a given force, the elastic deformation of the machine  $\Delta_2$  is a multiple of that of the specimen  $\Delta_1$ , a small inelastic deformation  $d\Delta_1$  will practically not affect the acting force, since the change of this force is determined by  $d\Delta_1/\Delta_2$ , which is very small. If, however, the elastic deformation of the machine is

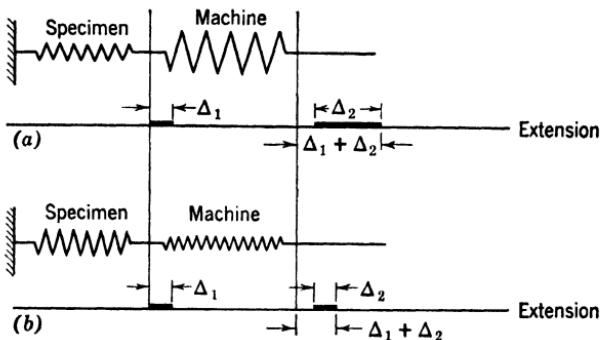


FIG. 85·1 Schematic representation of (a) "soft" and (b) "hard" deformation machine.

of the same order of magnitude as or even smaller than that of the specimen, the machine appears *hard*; the effect of the inelastic deformation  $d\Delta_1$  sharply reduces the acting force since  $d\Delta_1/\Delta_2$  is comparatively large. Hence, the *hardness* or *softness* of the machine is an important factor in the interpretation of the recorded force-deformation diagram, particularly since the slope of the unloading part of the diagram and the drop of the load at the yield point is largely determined by the character of the machine.

When the hard machine is stopped at the moment the specimen starts to flow, the load drops abruptly to a new equilibrium level determined by the relaxation of the elastic stress in the specimen resulting from its inelastic flow under the imposed constant deformation at which the machine was stopped. If the machine is not stopped but the over-all deformation of the specimen proceeds at a constant rate after the specimen has started to flow, the relation between deformation and load will depend on the

ratio between the rate of over-all deformation imposed by the machine and the rate of flow of the specimen. The smaller this ratio, that is, the smaller the imposed strain rate or the more rapid the flow, the more abrupt and the more pronounced the drop of the load at the moment of yielding. If the imposed strain rate attains or exceeds the rate of flow of the specimen, the deformation will proceed at constant or increasing resistance (load), the performance of the hard machine becoming identical with that of the soft machine. Hence, the hardness or softness of a testing machine is determined not only by the machine characteristic but also by the ratio of the spring constants of machine and specimen. Thus, for instance, a machine that appears soft in tests of a material at a certain strain rate and temperature may become hard if used for high-temperature testing of the same material.

Because of the heterogeneity of the initiation of plastic flow in metals, the deformation of the specimen will not proceed at the constant strain rate imposed by the machine, but at very rapidly varying rates, which depend on the momentary ratio between the rate of flow in the specimen and the rate of imposed deformation. Momentarily increasing flow rates are associated with a drop of the load, whereas a momentary reduction of the flow rate is marked by an increasing load. It is this interplay between flow rate and load that sharpens the discontinuities in the recorded load-deformation diagrams of hard machines at relatively low strain rates.

In soft machines the effect of the imposed strain rate is considerably less pronounced than in hard machines, since the discontinuities in the relation of load and deformation are compensated by the relatively large elastic deformations of the machine. Thus, the soft deformation machine has a characteristic similar to that of a load machine but has the advantage of higher dynamic sensitivity.

It is because of the relatively small moving masses that the dynamic sensitivity of the *deformation machines* is considerably higher than that of the *load machines*. Therefore these machines are able to record rapid deformation processes which are not recorded by load machines, because their inertia is too high. Thus, in recording inelastic deformation the hard machines respond very sharply by a drop of the load to even the smallest

inelastic deformation, whereas the soft machines tend to smooth out all fluctuations and to sustain the load at a nearly constant level. The expression of the heterogeneity of the initiation of plastic deformation in metals by the development of slip lines will therefore be intensified in the hard and smoothed out in the soft machine. Hence, the selection of the required type of response of the testing machine must depend on the purpose of the test.

The inertia effects in the machine are of particular importance in high-speed and impact testing. Because it is therefore extremely difficult to record directly load-deformation diagrams, the observations are usually limited to the recording of the total energy expended in deformation or fracture. But even in this case the energy dissipated during the process of transfer of the impact into the specimen by inertia effects and deformation of parts of the machine is so considerable as to make the interpretation of results rather difficult. In most high-velocity tests in which force-deformation curves have been reported, these curves were obtained from the actually recorded displacement-time diagrams by double differentiation with regard to time. The resulting acceleration-time diagram is easily converted into a force-time diagram if the mass producing the impact is known. The force-deformation diagram is then obtained by correlating the displacement-time and the force-time diagrams.

The difference between load machines (or *soft* deformation machines) and *hard* deformation machines is equally significant for the interpretation of results of tests under repeated load cycles as it is for single-stroke tests. If the amplitude of the load cycle is kept constant, the amplitude of the deformation cycle will change during the test. But the relation between load or stress at fracture and the number of load cycles sustained to fracture is not affected by this change. However, when the amplitude of the deformation cycle is kept constant under testing conditions under which the elastic deformation of the machine, that is, of the "spring" which enforces the deformation of the specimen, is small in comparison with the deformation of the specimen, the load amplitude drops with the start of inelastic deformation. The harder the machine or the more extensive the inelastic deformation at the imposed amplitude of deformation, the more pronounced the drop of the load amplitude. Thus, the actual

intensity of the load cycle produced by way of the imposed deformation cycle is variable, and the recorded relation between load or stress at fracture and the number of load cycles sustained to fracture is the less reliable, the more extensive the inelastic deformation during the test.

The most frequently used conventional mechanical tests are: (1) the hardness test, (2) the uniaxial tension test, (3) the impact test and (4) the repeated-load (fatigue) test. The significance of these tests will now be analyzed and related to the associated inelastic deformation.

## 86. The Hardness Test

The hardness test is the simplest of all mechanical tests to perform. By attributing a numerical value to the resistance of a material to indentation of a specified size or depth, encountered by an indenter of specific shape and assumedly infinite rigidity, it expresses a relation between the resistance in terms of the acting force, called hardness, and the inelastic deformation in terms of the size of indentation. In this definition *hardness* is the resistance associated with a certain arbitrarily specified amount of inelastic deformation and is therefore a function of this deformation. A less arbitrary procedure would be to define hardness in terms of *zero* inelastic deformation, as the maximum force under which no irrecoverable deformation, that is, no permanent indentation is produced. Based on this definition of hardness, however, the hardness test could hardly be performed.

Evidently the information which the hardness test provides will be related to the properties of the bulk of the material only if the properties of the surface do not differ considerably from those of the interior or if the thickness of the surface material of different properties is small in comparison with the depth of indentation.

In order to determine the resistance with which the material opposes the indentation, that is, its hardness, a large number of testing procedures have been developed. Of these the Brinell (ball indenter), the Rockwell (ball or cone indenter), and the Vickers (pyramid indenter) tests are the most widely used. In all these tests, the hardness is expressed either in terms of the indentation depth, if the applied force is invariable, or in terms of a force or of a uniform stress over the indentation area.<sup>86-1</sup>

The hardness values obtained from different tests are therefore less the expression of a genuine physical property than of comparative behavior under the different very specific testing conditions. Because of the complex state of stress produced in any type of hardness test, the hardness number must, in general, not be considered a real material constant. In Brinell tests it is highly variable and depends strongly on the absolute value of the indentation force; although in cone and pyramid indenter tests this dependence on force is very nearly eliminated because of the geometrical similarity of the indentations under any load, the hardness number itself is still a comparative characteristic rather than the expression of a physical property, since it depends on other factors of the test, such as the cone angle.

The definition and measurement of hardness in terms of a specific pressure alone cannot be considered adequate, since the significance of the resistance to inelastic deformation can only be interpreted in relation to this deformation. The assumption that an interpretation is possible in terms of the resistance alone would be valid only within the elastic range of deformations because, within this range, the form of the relation between resistance and deformation is known and independent of the test. The conventional procedure to represent the results of hardness tests in terms of the resistance to deformation alone is equivalent with the attempt to represent comparative results of tension tests in terms of the recorded loads or the true stresses at a single isolated and arbitrarily selected value of strain which is not specified. A real resistance value must, however, embody both cause and effect, that is, the applied specific force *and* the produced deformation. Hence, the real measure of resistance or hardness should be the ratio between the applied specific force (stress) and the resulting irrecoverable deformation (indentation). The resistance so defined would be more of a physical characteristic of the material than the conventionally defined hardness number; it would be of the type of a *hardening modulus*, similar to the concept of the secant modulus in tension tests beyond the elastic range. The fact that the surface over which the force is distributed is connected by a simple geometrical relation with the inelastic deformation (indentation depth) is a feature of the hardness test which makes measurements particularly simple, since, at least theoretically, it is sufficient to measure

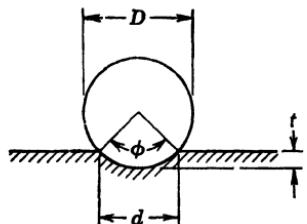
one of the geometrical characteristics only, that is, either diameter or depth of indentation.

Since the *hardening modulus* observed in hardness tests will, to a considerable extent, be determined by the work-hardening capacity of the material, which tends to decrease with increasing strain, a similar trend of decreasing hardening modulus with increasing indentation must be expected and has actually been observed. This similarity, however, is merely one of trend, since not only is the state of stress and therefore the extent of inelastic deformation different in both tests, but also the definition of deformation. In the hardness test this deformation cannot be expressed in dimensionless form (strain) by referring it to a definite measured length, but it is an average of a complex nonuniform deformation. Therefore the hardening modulus is still not a real material constant, but it is nearer to one than the conventional hardness number. Since the observed variables in the hardness tests are thus not independent of the geometry of the test, the correlation of the results of hardness tests with the results of tests in which mechanical properties are observed as functions of strain must necessarily remain empirical.

The two variables observed in the Brinell hardness test are the force  $P$  divided by the area  $A$  of the indentation, which is the hardness number,

$$H_B = P/A = P/\pi D t = \frac{2P}{\pi D(D - \sqrt{D^2 - d^2})} \quad (86 \cdot 1)$$

where  $D$  denotes the diameter of the ball and  $d$  the diameter of the indentation; the depth of the indentation  $t$  is a function of  $d$ :



$t = \frac{1}{2}(D - \sqrt{D^2 - d^2}) \quad (86 \cdot 2)$

In terms of the indentation angle  $\phi$  (Fig. 86·1):

$$H_B = \frac{2P}{\pi D^2(1 - \cos \phi/2)} \quad (86 \cdot 3)$$

FIG. 86·1 Brinell hardness test.

Hence, for similar indentations defined by equal angles  $\phi$  and identical hardness number  $H$  the loads are proportional to the squares of the ball diameters. Similar

formulas for hardness numbers  $H = P/A$  are obtained for indenters in the shape of cones or pyramids.

The relation between the force  $P$  and the diameter of indentation  $d$  is frequently presented in the form of a power law,

$$P = ad^n \quad (86 \cdot 4)$$

where  $a$  and  $n$  are constants that depend on hardness and diameter of the indenter ball,  $a$  representing the value of the load for the diameter  $d = 1$ ; the power  $n$  varies between  $n = 2$  and  $n = 3$ . For cone and pyramid indenters, for which the hardness number is independent of the load, the power  $n = 2$  and  $P = at^2$ .

The unsatisfactory correlation between the real hardness of a metal which is essentially the expression of its work-hardening capacity and the conventional hardness number, particularly the Brinell number, is evident from the shape of the observed  $H_B(P)$  and  $H_B(d)$  relations which, for a number of materials, show a definite maximum (Fig. 86·2). This maximum is unrelated to the steadily increasing work hardening of the metal but is a purely geometrical characteristic of the test. The hardening modulus  $H/t$ , on the other hand, shows a decline which, rapid at first, becomes more gradual later (Fig. 86·3); this behavior is consistent with the fact that the work-hardening modulus of a metal observed in a uniaxial test decreases at first rapidly and then more slowly and tends asymptotically towards zero. Correlation of hardness with mechanical characteristics observed in other types of tests depends on the similarity of the inelastic behavior under the compared testing condition.

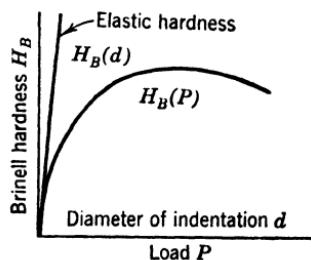


FIG. 86·2 Schematic relations of Brinell hardness and diameter of indentation or loads.

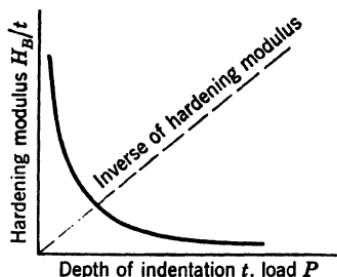


FIG. 86·3 Schematic relation of hardening modulus  $H_B/t$  and load or depth of indentation.

This maximum is unrelated to the steadily increasing work hardening of the metal but is a purely geometrical characteristic of the test. The hardening modulus  $H/t$ , on the other hand, shows a decline which, rapid at first, becomes more gradual later (Fig. 86·3); this behavior is consistent with the fact that the work-hardening modulus of a metal observed in a uniaxial test decreases at first rapidly and then more slowly and tends asymptotically towards zero. Correlation of hardness with mechanical characteristics observed in other types of tests depends on the similarity of the inelastic behavior under the compared testing condition.

Within the elastic range the elastic modulus may be considered a direct measure of hardness. Within the inelastic range the resistance to indentation depends on both the elastic and the inelastic (work-hardening) resistance. The smaller the indentation, that is, the smaller the load or the higher the hardness of the material, the more important is the elastic component of the resistance in comparison to the inelastic component, and the more nearly linear the relation between hardness and elastic modulus. This fact is an additional explanation of the difficulty of comparing hardness values associated with different depth of indentation.

No close empirical correlation between indentation hardness and yield limit observed in a tension test can be expected, since the amount of inelastic strain associated with the yield limit is considerably smaller than the inelastic strain produced by the indenter. It has been estimated that the inelastic strain associated with the usual Brinell hardness numbers of steel is of the order of magnitude of between 5 and 15 percent, whereas the yield limit is usually determined from the 0.2 percent offset. Because of the relatively large strain associated with the hardness test, a direct relation can be expected to exist between the hardness number and the stress  $s_t$  at which necking starts in the tensile test, since the localized strain at the early stage of necking is of a similar order of magnitude. The existence of such a relation between the hardness number  $H_u$  and the so-called *ultimate tensile strength*  $s_t$  (see Art. 87),

$$s_t = CH_u \quad (86 \cdot 5)$$

is actually observed;<sup>86·2</sup> the factor  $C$  varies between 0.39 and 0.45. A relation of this type can be expected to hold as long as the order of magnitude of inelastic strain is comparable in both types of test, and to become the more unreliable, the more dissimilar, for a certain material, the inelastic aspect of the respective tests.

The significance of time and of temperature in the interpretation of the results of hardness tests depends on the deformational character of the material. In hardness tests of essentially viscoelastic materials, both the rate of indentation and the time of load application, and also the period elapsing between load removal and the measurement of the indentation, will necessarily

affect the result, since the influence of creep and elastic after-effect (creep recovery) may be considerable. The results of indentation hardness tests of metals at elevated temperatures may be affected both by the change in hardness with temperature of the indenter, by the difference in temperature between the indenter and the material, as well as by the duration of the test and the period between load removal and measurement of the indentation. In order to obtain reproducible results all those factors must be controlled.

Indentation hardness tests are frequently used in order to control the uniformity of a certain property related to hardness, such as for instance the conventional "ultimate tensile strength." If the correlation is satisfactory and interpretable in terms of the principles of inelastic behavior, as it is in the case of tensile strength according to eq. 86·5, hardness tests may, for control purposes, be preferable to tension tests because of their simplicity. The statistical information concerning the uniformity of the "ultimate tensile strength" of the material obtainable from a large number of simple and rapidly performed hardness tests is usually more relevant to the purpose of the test than the limited information supplied by a small number of elaborate tension tests.

The correlation of data concerning performance in technological processes with conventional hardness numbers is rather vague. A better correlation may however be obtained by using the hardening modulus. With regard to design, the hardness number is not directly significant, since, like the "ultimate tensile strength," it is associated with inelastic deformation far beyond that involved under service conditions.

Apart from hardness measurements by indentation under static loads, dynamical testing procedures have been developed in *scleroscope tests* in which the hardness is measured in terms of the height of rebound  $h$  of a body that is dropped from a specified height  $h_0$  on the surface of the material to be tested. It is evident, that the concept of hardness as defined by the rebound test is basically different from the hardness concept underlying the indentation test. Whereas, in the indentation test, hardness is defined as the resistance to inelastic deformation, it is defined in the rebound test as a direct function of the recoverable potential energy, which is the energy that produces the rebound.

Hence, the relation measured in the scleroscope test is that between the recoverable strain energy and the applied energy. The rebound is therefore related to the damping capacity of the material rather than to its hardness.

Even for the purpose of comparative hardness testing, the rebound test can evidently be applied only to materials of equal modulus of elasticity. Only in this case is the trend of the rebound values related to the trend of hardness values, since, the higher the rebound, the smaller the amount of energy spent in producing inelastic deformation; the smaller therefore the inelastic deformation itself, which, in terms of the indentation test, is inversely proportional to the hardness of the material under constant load. Repetition of the rebound test at the same spot must necessarily result in increasing *rebound hardness* values, because of the work-hardening effect produced by the test itself.

The results of rebound tests of different materials are comparable only if they are interpreted in terms of damping and not of hardness. The specific damping may be assumed to be roughly related to  $h_0$  and  $h$  by the equation,

$$\psi \sim \frac{h_0 - h}{h_0} = 1 - h/h_0 \quad (86\cdot 6)$$

from which the other damping constants can be derived.

## 87. The Tension Test

The tension test is the basic mechanical test. It is at present the most extensively used method for evaluating the physical properties of solid and apparently solid materials. At the same time the tension test is a very complex procedure; its results depend to a considerable extent on the geometrical form of the specimen, the characteristics of the testing apparatus, and the general conditions of the test.

The results of tension tests are recorded either in the form of load-deformation curves or of stress-strain diagrams. Whereas the load-deformation curve is easily defined and recorded, its use in the interpretation of the test is unsatisfactory, since it depends essentially on the dimensions of the test specimen. On

the other hand, the use of diagrams of stress versus strain, which is more satisfactory because of the recording of variables which are essentially independent of the particular dimensions of the test specimen, introduces the difficulties of an adequate and consistent definition and measurement of those variables. This is but a particular aspect of the quite general experience in the mechanical testing of materials that, the easier and more rapidly a test can be performed, the more difficult the interpretation of the observed results. The difficulty of defining and recording stresses is a characteristic of the tension test and due to the nonuniformity of the deformation process within the inelastic range of the test; the difficulty and arbitrariness of defining strain is characteristic for the analysis of all deformation processes in which deformations are not infinitesimal (see Art. 29).

Stress in the tension test is defined either as a fictitious value, by dividing the applied load by the original area of the cross section of the specimen (nominal or "engineering" stress  $s_0$ ) or as a real value by dividing the applied load by the smallest real area pertaining to this load ("true" stress). This value is, however, real only within the range of uniform strain, becoming fictitious after necking sets in and the state of stress is three-dimensional.

The strain is derived either from the extension of a certain gage length  $l_0$  or from the reduction of the original cross-sectional area  $A_0$ . A definite geometrical relation between the changes of  $l_0$  and of  $A_0$  exists as long as the deformation is homogeneous. Beyond this range, that is, when necking sets in, the changes of length and diameter can no longer be related; strains derived from the observed over-all extension and from local area reduction are therefore no longer interchangeable. Since the measured extension is not uniform over the gage length but represents the sum of a uniform elongation and a highly localized nonuniform deformation, the strain derived from the maximum area reduction is a more adequate measure of inelastic deformation within the range of necking than that derived from the over-all extension of a gage length.

The principal definitions of strain which are used in representing the results of tensile tests are<sup>87-1</sup> (for assumedly volume-constant materials):

(a) The nominal elongation strain  $e_0$ ,

$$e_0 = \frac{\Delta l_0}{l_0} = \frac{l - l_0}{l_0} \quad (87 \cdot 1)$$

or, since  $Al = A_0l_0$ ,

$$e_0 = \frac{A_0 - A}{A} = \frac{\Delta A_0}{A} = q_1 \quad (87 \cdot 2)$$

(b) The nominal area-reduction strain  $q_0$ ,

$$q_0 = \frac{A_0 - A}{A_0} = \frac{\Delta A_0}{A_0} \quad (87 \cdot 3)$$

or, because of volume constancy,

$$q_0 = \frac{l - l_0}{l} = \frac{\Delta l_0}{l} = e_1 \quad (87 \cdot 4)$$

(c) The logarithmic (natural) strain,

$$e = \int_{l_0}^l \frac{dl}{l} = \log\left(\frac{l}{l_0}\right) = - \int_{A_0}^A \frac{dA}{A} = \log\left(\frac{A_0}{A}\right) = q \quad (87 \cdot 5)$$

The relation between the various strains are

$$\begin{aligned} e = q &= -\log(1 - q_0) = -\log(1 - e_1) = \log(1 + e_0) \\ &= \log(1 + q_1) \end{aligned} \quad (87 \cdot 6)$$

and

$$1 + e_0 = \frac{1}{1 - e_1}; \quad 1 + q_1 = \frac{1}{1 - q_0} \quad (87 \cdot 7)$$

For different values ( $A_0/A$ ) or ( $l/l_0$ ) the functions  $e_0 = q_1$ ,  $q_0 = e_1$  and  $e = q$  have been plotted in Fig. 87·1.

The complexity of the behavior of metals in the tensile test is due to the fact that, beyond the elastic range, the deformation is not uniform but, because of the heterogeneity of the flow, proceeds in a series of local extensions associated with "migrating" small local contractions.<sup>87·2</sup> At each point of contraction the stress increases; however, at the same section the material has also been work-hardened. There is therefore a competition between the increase of resistance to inelastic deformation due

to work hardening and the increase of stress due to area reduction. As long as the work hardening overcompensates the area reduction, the inelastic straining is transferred into another less work-hardened section. However, as soon as the rate of work hardening has been so far reduced that it can no longer compensate the local reduction of area, the section appears weaker than the adjacent sections and the deformation continues in the same section, in which therefore the stresses increase and necking sets in. Necking is thus an instability condition defined by a maximum of the load  $P$  that the affected section is able to carry. Since  $P = sA$ , the condition at which necking starts can be expressed by  $dP = d(sA) = 0$  and therefore (see Art. 45):

$$\frac{ds}{s} = - \frac{dA}{A} \quad (87 \cdot 8)$$

As long as the increase of the total force  $A ds$  due to work hardening exceeds the decrease ( $-s dA$ ) due to area reduction, the carrying capacity of the section

increases, and "uniform" inelastic elongation is produced by propagation of the local plastic strain from section to section. As the rate of the purely geometrical area reduction cannot be expected to increase discontinuously, the fact that, at a certain strain, the rate of work hardening is so substantially reduced that  $A ds < s dA$ , could be the result of the rather rapid change in the work-hardening mechanism due to a change in the hardening effect of crystal fragmentation after the volume of crystals of largest size has been broken up, as discussed in Art. 45, or it could be the result of a sharp intensification of recrystallization.

Since the instability of the deformation of the specimen in the tension test thus appears to be the expression of a certain change in the work-hardening process, the stress at which this change occurs has no significance beyond that of delimiting two stages of the work-hardening process. The conventional interpretation in terms of a characteristic of strength (*ultimate tensile strength*)

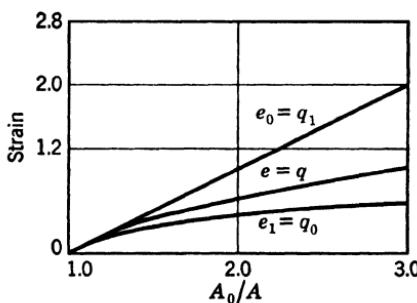


FIG. 87·1 Comparison of various definitions of strain in terms of  $A_0/A$  (after McGregor<sup>87·1</sup>).

of the maximum load or of the nominal stress at which necking sets in has therefore no physical justification; hence, the recorded value of the conventional tensile strength does not represent a characteristic of the material interpretable in terms of structural performance or significant in design. This value is conspicuous only if the results of the tension test are presented in terms of the load or the nominal stress  $s_0$  but loses its identity as a maximum if the recorded dynamical variable is the true stress  $s$ . In this representation it is, however, defined by a certain characteristic relation between stress and strain.

If the recorded strain is the nominal strain  $e_0$ , the point of maximum load in the true stress versus strain curve is easily determined as the point at which the tangent through  $(e_0 = -1, s = 0)$  touches this curve. For a different definition of strain the identification is less simple. The condition for the point of maximum load follows from eq. 87·8 if  $A$  and  $dA$  are expressed in terms of the strain, by introducing the assumption of volume-constant deformation. Since, according to eqs. 87·1 to 87·5,

$$-\frac{dA}{A} = \frac{de_0}{1 + e_0} = \frac{de_1}{1 - e_1} = de \quad (87\cdot9)$$

the following relations are obtained from eq. 87·8 for the slope of the tangent at the point of maximum load (Fig. 87·4).

$$\frac{ds}{de_0} = \frac{s}{1 + e_0}; \quad \frac{ds}{de_1} = \frac{ds}{dq_0} = \frac{s}{1 - e_1} = \frac{s}{1 - q_0}; \quad \frac{ds}{de} = s \quad (87\cdot10)$$

It is frequently assumed that the true stress—logarithmic-strain relation between the yield stress and the stress at maximum load (necking stress) can be represented by a simple power function, because of the relatively small range over which this relation has to be so represented; such assumption is generally justified, although it should be kept in mind that the proposed representation is a purely empirical curve-fitting procedure. Hence, if

$$s = \text{const } e^n \quad (87\cdot11)$$

reproduces the recorded true stress-strain diagram within the considered range, the double-logarithmic representation of this function is a straight line. Because

$$\frac{ds}{de} = n \text{ const } e^{(n-1)} \quad (87 \cdot 12)$$

and because of eqs. 87·8 and 87·9, the relation is obtained,

$$\text{const } e^n = n \text{ const } e^{(n-1)} \quad (87 \cdot 13)$$

which shows that the logarithmic strain at the maximum load is expressed by the power  $n$ . Since according to eq. 30·11 this power  $n$  represents the slope of the function 87·11 in double-logarithmic representation, this slope gives the logarithmic strain at the maximum load. The value of this strain can thus be fairly accurately determined from a double-logarithmic plot of the true stress-strain diagram, however, only as long as the stress-strain diagram can be represented by a single power function.

In actual tests there is usually a slight difference between the points on the stress-strain diagram at which the maximum load is reached and the point at which necking sets in, since necking starts somewhat before the maximum load is attained. This difference is the result of the time effects which have not been considered in the analysis of the tension test. Since the apparent resistance of the specimen to deformation is not only a function of work hardening, but also a function of the strain rate, the insetting necking which tends to increase the local strain rate raises the stress at which the deformation proceeds beyond the work-hardening resistance. As a result the maximum load is slightly higher than the necking load, the difference being the larger, the more pronounced the intrinsic effect of the strain rate.

In Fig. 87·2 the results of a tension test have been plotted in terms of both nominal and true stresses, as well as for the three different definitions of strain in terms of area reduction  $q_1$ ,  $q_0$ , and  $q$ . Although both the functions  $s = f(q)$  and  $s = f(q_1)$  are very nearly straight lines beyond the point of maximum load, the function  $s = f(q_0)$  shows a marked upward curvature. This upward curvature, which is observed in practically all tension tests the results of which are plotted in the variables  $s = f(q_0)$ , is the result of the particular definition of the strain. The straight-line part of the relation  $s = f(q)$  or  $s = f(e_0)$ , on the other hand, has no particular significance; it extends only over a certain range of strains, beyond which it may be expected to curve down, unless fracture intervenes (see Art. 45).

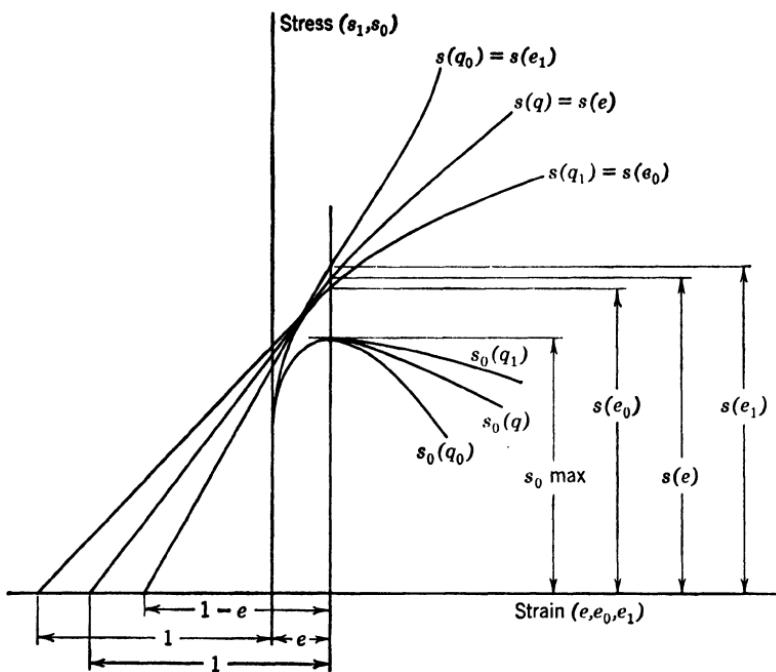


FIG. 87-2 Comparison of different representation of results of tension test of mild steel specimen (after McGregor<sup>87-1</sup>).

The representation  $s = f(q_0)$  has been used to extrapolate the results of tensile tests to values of  $q_0 = 1$  or  $A = 0$ , defining the extrapolated stress  $s$  at  $q_0 = 1$  as the *true cohesive strength*<sup>87-3</sup> (Fig. 87-3). This procedure tacitly neglects the characteristic upward curvature of all  $s = f(q_0)$  diagrams for large values of strain by replacing the actual curved diagram by a tangent at the maximum load stress, a procedure for which no justification exists. If the abscissa  $q_0 = 1$  is interpreted in terms of  $q$ , which is the only definition of strain that does not lose its meaning for very large strains, it is found that, for  $q_0 = 1$ , the strain  $q = \infty$ , whereas, for  $q_0 = 0.67$ , the pertaining value  $q = 1.10$ . Hence, the range of strains in the representation  $s = f(q)$  between

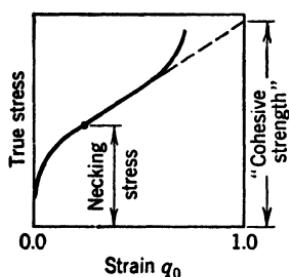


FIG. 87-3 Ludwik's extrapolation of the  $s(q_0)$  curve.

$q = 1.10$  and  $q = \infty$  is shortened and thus sharply distorted in the representation  $s = f(q_0)$  to a range between  $q_0 = 0.67$  and  $q_0 = 1.0$  (Fig. 87·4). Evidently, the distortion of the function  $s = f(q)$  in the representation  $s = f(q_0)$  becomes the more pronounced, the nearer  $q_0$  approaches unity; the upward slope of the  $s = f(q_0)$  diagram may thus be expected to become very steep in the vicinity of  $q_0 = 1$ . The limiting ordinate of  $s$  at  $q_0 = 1$  which represents the limiting work-hardening resistance  $s_\infty$  for  $q = \infty$ , and which is reached asymptotically (see Art. 45) can

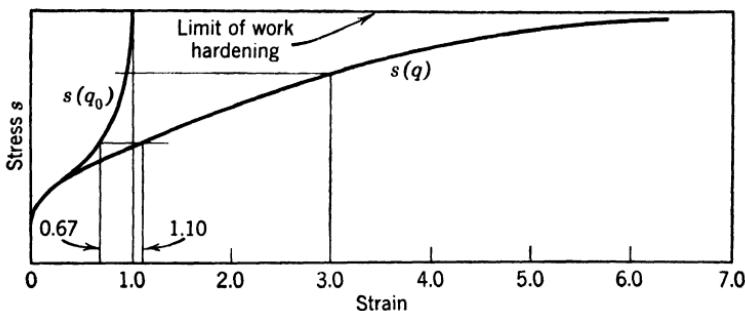


FIG. 87·4 Comparison of representation  $s(q_0)$  and  $s(q)$  for large strains.

therefore have no relation whatsoever to a fictitious value of stress obtained by linear extrapolation.

The same is true if the test results are plotted, as has been done by Stead and others,<sup>87·3</sup> in terms of  $s = f(d/d_0) = f\sqrt{1 - q_0}$  instead of  $s = f(q_0)$  (Fig. 87·5). Although, in this representation the upward curvature of the diagram starts at considerably higher values of "strain," the range of deformation  $0.33d_0 \geq d \geq 0$  is, in terms of logarithmic strain, the range  $2.3 \leq q \leq \infty$ . The work-hardening function in terms of  $\psi = f(q)$  extending from  $q = 2.3$  to infinity is therefore distorted beyond recognition by being squeezed into the range  $0.33d_0 \geq d \geq 0$ . In the vicinity of  $d = 0$  an upward curvature of the  $s = f(d/d_0)$  diagram of such sharpness must therefore be expected that any attempt to approximate this diagram by a straight line is clearly invalidated. Even an extrapolation from  $q_0 = 0.999$  to  $q_0 = 1.000$  is impossible since, for  $q_0 = 0.999$ , the logarithmic strain  $q = 3.0$ ; thus the stress-strain diagram  $s = f(q)$  between  $3 < q < \infty$  would, in the  $s = f(q_0)$  representation, be squeezed into the interval  $0.999 < q_0 < 1.0$ .

The stress-strain diagrams  $s = f(q_0)$  representing the results of tension tests of the same material at different stages of work hardening (Fig. 87·5) must necessarily intersect at  $q_0 = 1$ , since in the particular definition of strain  $q_0$  this intersection is the expression of the fact that an ultimate work-hardening limit exists which is reached asymptotically, that is, for  $q = \infty$ . The

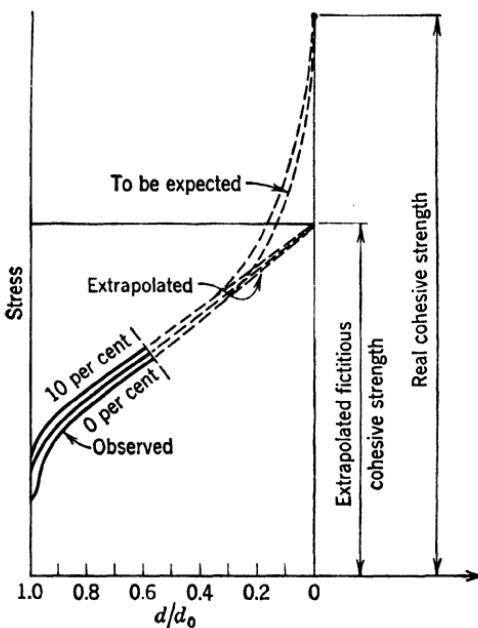


FIG. 87·5 Extrapolation of  $s(d/d_0)$  curves recorded in tension tests of carbon steel at different stages of work hardening defined by the percentage of area reduction at the end of the work-hardening process (after Stead<sup>87·3</sup>).

point of intersection can, however, not be found by linear extrapolation from strains  $q_0 < 1.0$ .

The information contained in the stress-strain diagram is the level of yield stress, of necking stress, and of fracture stress, and the work-hardening capacity of the material as indicated by the slope of the diagram between the necking stress and the fracture stress. Only the first two figures are obtained from the conventional nominal stress versus strain diagram. Of these the yield stress is the only figure of direct engineering significance. It

determines the load-carrying capacity of all structures for which dynamical effect are not of primary importance. The less pronounced the yield stress, the less clear its engineering significance, since the carrying capacity of a structure can be clearly defined in terms of the yield stress only if yielding is a real rather abrupt transition of deformational behavior of the material (see Art. 79). If the yield stress is a designated limit, it is the shape of the transition from the elastic into the work-hardening range and the slope within the work-hardening range that become of significance in the evaluation of the material.

The shape of the true stress-strain diagram within the transition range between the designated yield stress and the necking stress is of significance as an indication of the ability of the material to redistribute elastic peak stresses by inelastic action. The more gradual the transition, the less relief of stress concentrations can be expected and the more liable therefore the material to brittle fracture. On the other hand, the ability of such materials to sustain a stress exceeding the designated yield stress without appreciable deformation associated with slip and crystal fragmentation reduces the damage within the internal structure produced by this *overstress* and makes the material resistant to its repetition. Thus metals with a rather sharp transition into the work-hardening range may be expected to be less sensitive to stress concentrations, but more sensitive to repeated overstress, than metals with a gradual transition.

For metals of roughly equal fracture stress the slope of the diagram between the necking stress and the fracture stress is significant with regard to the performance in technological and fabrication processes rather than in the finished structure. It is an indication of the forces required to produce deformation in deep-drawing, stamping, cold-rolling, and other cold-forming processes, as well as of the suitability of the material for large local deformation without embrittlement. This correlation of test and performance is possible, since the respective strains are of the same order of magnitude. However, the slope of the stress-strain diagram alone is not a sufficient guide to ductility, if the fracture stresses of the compared metals are different. In this case the total amount of dissipated energy, as defined by the area under the stress-strain diagram, in conjunction with the slope within the work-hardening range, might provide a more

reliable indication of cold deformability. The best materials would be those with the largest areas and the smallest slopes.

Although the structural performance of a metal is strongly affected by the shape of the  $s = f(q)$  diagram in the vicinity of the yield stress, its suitability for actual use in structures depends to a considerable extent on the ratio between the yield stress and the necking stress. This figure has, however, no relation to the actual performance of the structure under any possible service condition, but only to its expected performance under abnormal accidental conditions. Both the yield stress and the necking stress are limits of inelastic behavior; the range between those two stresses is characterized by a work-hardening capacity which is considerably higher than that beyond this range. Hence, it is only within that range that a deformational resistance, appreciably exceeding the resistance at the yield stress, can be developed by inelastic deformation; the strains, although relatively large, are still limited by the existence of states of equilibrium. The consequences of an accidental excess load, particularly one that is dynamically applied, are rendered less serious if the progress of yielding is accompanied by an increasing resistance by which it can be checked before the deformation is excessive or fracture occurs. Evidently, the amount of energy that can be absorbed in this process depends as much on the difference between the two stress levels as on the necking strain. It is therefore the ratio between yield stress and necking stress in conjunction with the limit of uniform strain, that is, the strain at insetting necking, that provides the reserve in carrying capacity against accidental, principally dynamic, overloads, ensuring that the work-hardening capacity mobilized against the acting forces be adequate but also that the uniform yielding be sufficient to *reduce* the intensity of the forces resulting from the impact. That the "strength reserve" assumedly represented by the excess of the necking stress over the yield stress is unrelated to the design and does not define an ultimate carrying capacity, which could be compared to a design-carrying capacity, is evident from the fact that the *ultimate tensile strength*, as the necking stress is usually called, is neither "ultimate" nor a "strength," but only a point at which the rate of work-hardening resistance changes; it is situated within a range where strains are a multiple of maximum design strains.

There is no relation between the information obtained from a uniaxial tension test and the performance of the material under conditions of triaxial stress. High ductility in the tension test has no bearing on the actual behavior under triaxial stress, particularly as the strain values associated with ductility in the uniaxial test are of a different order of magnitude from the strains under conditions of triaxial stresses. Satisfactory performance under triaxial stress requires a relatively low yield limit in terms of distortional energy, so that, even under conditions for which the ratio between volumetric strain energy and total applied strain energy is high, a relatively small amount of distortional energy is sufficient to produce inelastic relief of peak stresses. Thus a stress-strain diagram starting to curve downward at a relatively low stress level might indicate a better performance of the metal under conditions of triaxial stress than a linear relation attaining at high yield stress.

The true fracture stress observed in a uniaxial tension test represents the cohesive strength, under the particular state of stress under which fracture occurs, of a material that has undergone the inelastic deformation preceding fracture. These conditions are, however, unknown, unless the specimen fractures without appreciable inelastic deformation. The fracture stress of the deformed material can therefore have no significance with respect to the performance of the material in service, since even the most critical service conditions are not associated with inelastic deformations of the order of magnitude observed at fracture in the uniaxial tension test of a ductile material. The fracture stress can be directly related to the carrying capacity of the structure only if both the specimen and the structure fail in brittle fracture. However, the smaller the inelastic deformation that precedes fracture, the more important the local inhomogeneities of stress both within the structure and within the specimen, and the less reproducible therefore the results of tests. The greater therefore the difficulty of establishing a correlation between the results of similar individual tests and between test results and performance.

Strain at fracture, either in terms of elongation or of area reduction, is frequently considered a measure of the ductility or the toughness of the material. Actually these values are functionally unrelated to the performance of the material under

service conditions, since the necking of the tensile specimen, which is the principal factor in producing the strain values at fracture, is a characteristic instability effect of the test under conditions not comparable to service conditions. Because of the necking, however, a triaxial state of stress develops in the vicinity of the necked-down section. Thus, the specimen at stresses beyond the necking stress is subject to a notch-tension test, and its performance in this test is an indication, at least qualitatively, of its ductility under a hydrostatic stress component superimposed on the distortional stress, which increases with the reduction of the necked-down section. Hence, it may be assumed that the more extensive the necking prior to fracture, the better the performance of the material around a notch or a stress concentration. The measure of such ductility, however, is not the over-all elongation of the gage length at fracture, but the local reduction of area of the necked-down section.

Whereas the area reduction at fracture may thus be considered a comparative measure of notch ductility, the practically uniform elongation associated with the necking stress is an indication of the maximum capacity of the material to absorb, that is, to dissipate energy under exceptionally heavy excess loads with a limited amount of deformation. Since this capacity is unrelated to the local necking, it is also not clearly related to the over-all elongation at fracture. Being a combined expression of the uniform elongation prior to necking and of the local elongation associated with the local contraction, the over-all elongation at fracture is an unsatisfactory characteristic of ductility, as in the ductility evaluation of the material it is desirable not to combine the uniform and the local inelastic deformation.

The magnitude of inelastic strain necessary for redistribution of stress concentrations and relief of elastic peak stresses is, in general, of the order of magnitude of the elastic strains (see Art. 43). Thus, satisfactory performance of a material under service conditions would require not more than this amount of inelastic strain accompanying stresses of the order of magnitude of the highest elastic service stresses. The inelastic strain at fracture or even at the necking stress is therefore irrelevant with regard to service performance, unless it is assumed that in a continuous stress-strain diagram an inelastic strain of the order of magnitude of the elastic strain at the level of the highest service stress is

necessarily associated with a definite limit of uniform inelastic strain prior to necking. This limit of uniform strain would thus not be significant in itself, but only as an indication of the magnitude of inelastic strain under extreme service conditions.

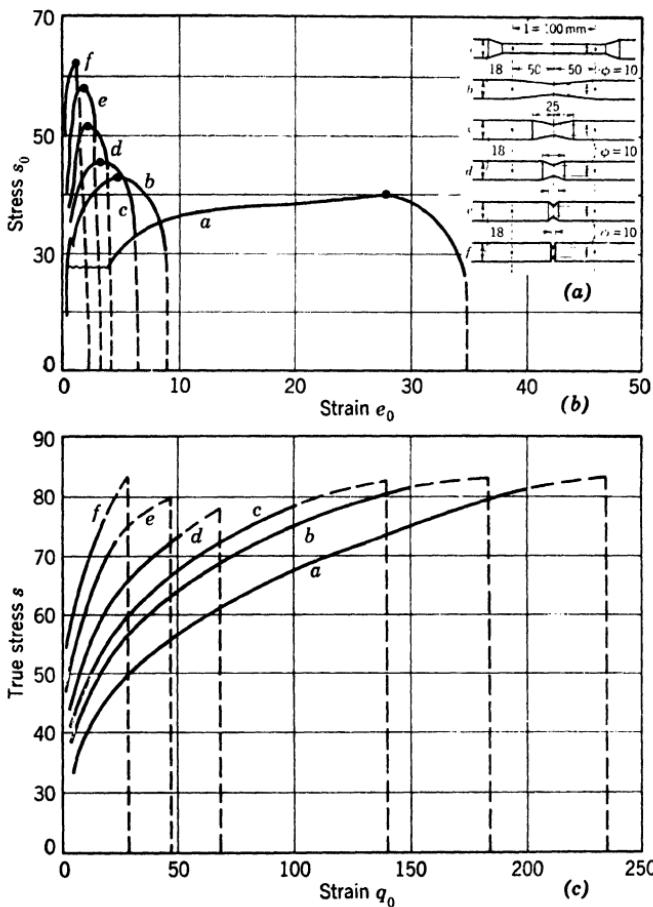


FIG. 87.6 Comparison of effect of different notches (a) on stress-strain diagrams in tension test, using (b) nominal stresses, or (c) true stresses (after Ludwik<sup>87.3</sup>).

Conventional creep tests are tension tests under sustained load. Because of the relatively small strains involved, the difference in the various definitions of strain are of less importance in creep tests than in the conventional tension tests. Similarly, the

consideration of the true stress instead of the load or the nominal stress does not appreciably affect the results. For large strains, however, it may be necessary not only to correct the results recorded in terms of nominal stress  $s_0$  and strain  $e_0$ , but also to compensate for the increase in stress due to the reduction of cross-sectional area, since creep tests should actually be performed under constant stress (see Art. 49).

If the tensile test specimen contains a symmetrical notch, the distortional energy in the vicinity of the notch, resulting from the applied axial load, will be the smaller the higher the specific volumetric expansion due to the notch. Hence, the load or stress required to produce the same amount of inelastic strain increases with increasing sharpness of the notch, with the result that both the uniaxial nominal yield-stress and the ultimate stress are raised by notches of increasing sharpness (Fig. 87·6). The fracture stress itself appears to be unaffected by the character of the notch; however the extent of work hardening required to bring the stress up to the critical value, which is indicated by the amount of irrecoverable deformation preceding fracture, necessarily depends on the form of the notch; it is this form which, by determining the relative amounts of volumetric and distortional energy associated with a certain applied load, determines the amount of additional resistance to deformation to be provided by work hardening, before the total resistance attains the critical energy level characteristic of the fracture stress. Thus the higher the stress that can be sustained within the notched section without inelastic deformation, the smaller the inelastic strain at which the stress reaches the limit of fracture.

## 88. The Torsion Test

Among mechanical tests the torsion test is theoretically the most adequate test for studying inelastic behavior, since it represents conditions of purely deviatoric stresses. Therefore the torque-twist diagram observed in the test of the specimen has the same character as the basic shear-stress-shear-strain diagram for the volume element; both diagrams are directly proportional, if tests are performed on thin-walled cylinders in which the effect of the nonuniform stress distribution along the diameter of the plain cylindrical specimens is practically eliminated. Such proportionality does not exist in other but torsion

tests, since the isotropic stress superimposed on the deviatoric stress under conditions of simple tension, compression, or bending affects the force-displacement diagram without affecting the inelastic component of the deformation.

Like tension tests, torsion tests can be performed at constant strain rate or at constant stress rate. In the first case the torque is applied by a gear or by a transmission; in the second it is produced by a driving pulley from which a bucket is suspended into which lead shot or water is fed at constant speed. If this speed is kept very low, it is relatively easy to observe the discontinuous steplike types of stress-strain diagrams discussed in Art. 46 since the slip responsible for the discontinuity is not influenced by a volumetric component of strain.<sup>88-1</sup>

Because of the elimination of the volumetric expansion and of the necking associated with the inelastic deformation in the uniaxial tension test the range of inelastic strain that can be investigated in a torsion test considerably exceeds that produced in a tension test prior to fracture. This difference increases with increasing brittleness of the material. Thus, a previously cold-worked metal appears prior to fracture considerably stronger and less ductile in the tension test than in the torsion test,<sup>88-2</sup> a phenomenon that is probably due to the different effect of the density decrease associated with cold working; deformation in the tension test is accompanied by dilatation, whereas in the torsion test it proceeds at constant volume.

The relations between the mechanical variables under conditions of uniaxial stress and of torsion or pure shear have been given in Arts. 24 and 25. From the equations derived there it follows that the stress in the tensile test is in first approximation equivalent to double the shear stress in the torsion test, whereas the tensile strain is equivalent to one-half the shear (torsion) strain.

## 89. Impact Tests

The conventional designation of a test as an *impact test* is rather inadequate, since it is current practice to lump together, under the general heading of impact tests, tests on notched specimens usually performed in bending under impact, with high-velocity tension tests of unnotched bars, without considering the basic difference between these two types of tests. In the

notched-bar test the inelastic behavior of the material is mainly governed by the triaxial state of stress imposed by the notch and only to a lesser extent by the velocity of the striking mass, whereas in the dynamic tension test of unnotched specimens the effect of the strain velocity is of primary importance. In both types of tests, which are now carried out mainly on pendulum-type machines, the velocity of strain is assumed to be equal to the pendulum velocity. This assumption is justified only if the energy of the striking pendulum considerably exceeds the energy required to deform and break the specimen.

The purpose of either type of test is different. Dynamic tension tests are usually performed with the aim of observing the deformational behavior of a material strained at high velocity; notched-bar impact tests, on the other hand, are used for an empirical evaluation of the service performance of different material under dynamic loads. Whereas it is difficult to *perform* a dynamic tension test in such a way that the stress-strain diagram is reliably recorded, but not more difficult to *interpret* it than it is to interpret the static tension test, it is relatively easy to *perform* a notched-bar impact test and record the energy required to break the specimen, the so-called *impact value*, but extremely difficult to *interpret* the significance of this energy value in terms of basic mechanical properties of the material. This difficulty as well as the difficulty of comparing the results of test on specimens of different shape and size is due essentially to the difference in the inelastic deformation associated with the behavior under identical testing conditions of those specimens. However, the ease with which the conventional notched-bar impact tests, such as the Charpy or the Izod test, are performed has made these tests extremely popular.

The characteristic shape of the force-elongation diagrams recorded in dynamic tension tests expresses the velocity effects on the deformation to be expected on the basis of the analysis of inelastic behavior. With increasing strain rate the yield stress is raised above the level observed in the static tension test, approximately in accordance with eq. 19-13. Evidently this tendency will be the more pronounced, the more time-sensitive the inelastic behavior of the material is. Thus the yield stress of a cold-worked metal or the part of the force-elongation diagram that represents the transition from elastic to large-scale

inelastic deformation will be raised more pronouncedly in a high-velocity tension test than that of a material with a more stable and therefore less time- and temperature-sensitive structure, such as an annealed metal.<sup>89-1</sup> On the other hand, the *slope* of the dynamic force-elongation diagram is smaller than that determined in the static test. Again, this phenomenon, which is due to the thermal softening of the material by the heat energy dissipated in inelastic deformation and not carried away producing the nearly adiabatic character of the high-velocity test (see Art. 54), is the more pronounced, the larger the specific amount of dissipated energy, that is, the higher the velocity of the test, and the less stable, that is, the more temperature-sensitive the structure of the material. Thus, cold-worked metals will show a more pronounced decrease of the slope of the force-elongation curve in dynamic tests than metals in annealed condition. Because of the reduced work-hardening rate in the adiabatic test, the instability inherent in the tension test and expressed by eq. 87·8 is delayed, and necking sets in at a higher uniform strain. Hence, because of the possibility of a simultaneous increase of both stress and deformation, the total energy dissipated prior to fracture and considered indicative of dynamic performance may be higher in the dynamic tension test than it is in the static test. This difference will usually be more marked for cold-worked metals than for annealed metals and may practically vanish if the structure of the material is very nearly stable and the testing conditions more isothermal than adiabatic. Thus the deformational behavior in a dynamic test will depend on the ratio of volume to surface and therefore on the diameter of the specimen.

The foregoing conclusions hold only as long as the imposed strain rate of the test does not attain the *transition velocity*, at which fracture, preceded by large inelastic deformation, is transformed into a brittle-type fracture (see Art. 57), or at which the imposed strain rate exceeds the velocity of propagation of the plastic deformation wave (see Art. 54). For unnotched specimens, however, the limiting rates are much higher than the velocities attainable in the conventional tension impact machines. Even so, the interpretation of results of impact tests requires careful consideration of all aspects of inelastic behavior during the test. Otherwise, differences in the test results will be ascribed

to different properties of the material or to inadequate performance of the tests, where the differences may be due entirely to the different inelastic behavior resulting from differences in the testing conditions.

The impossibility of reaching the *transition velocity* of strain in tests under homogeneous states of stress has been one of the principal reasons for the development of the notched-bar impact test. The effect of a notch in a specimen in the dynamic test is twofold: It induces three-axial states of stress and thus a state of hydrostatic tension which, superimposed on the deviatoric stress, considerably affects both inelastic behavior and fracture; and it increases the specific rate of energy application by reducing the volume of the deformed material. Hence the total energy required to deform and break the notched specimen is a fraction of the energy required to break the unnotched specimen and becomes a function of the shape and dimension of the notch and of the dimensions of the specimen. For reasons of convenience the notched-bar impact test is usually performed in bending, which considerably simplifies the testing procedures but increases the complexity of the testing conditions and thus makes interpretation of test results more difficult.

No interpretation of the results of notched-bar impact tests in basic physical terms is possible, since these results depend on the geometrical features of the test. It can therefore not be reliably decided whether observed variations are the result of the differences in one or a number of geometrical factors, or whether they are the expression of differences in some intrinsic property of the material. The energy absorbed in breaking the specimen, recorded as the characteristic of the material in the test, expresses both the energy of deformation, elastic and inelastic, prior to the initiation of a crack at the root of the notch and the energy expended in the propagation of the crack across the specimen. As long as the fracture is preceded by appreciable inelastic deformation, the energy of crack propagation is probably very small in comparison with the energy of deformation. If, however, the energy of deformation is very small, as in brittle fracture, the share of the energy of crack propagation in the total energy may become considerable. Thus the total energy will become the nearer proportional to the width of the specimen across which the crack spreads, the more brittle the fracture becomes.

By recording the total amount of energy dissipated in ductile fracture, the notched-bar impact test does not distinguish between metals with a very high work-hardening rate and small deformation prior to fracture and those with a low work-hardening rate and large deformation. It combines, moreover, the effects of strain velocity and the state of triaxial tension at the root of the notch. It is therefore obvious that impact values which vary with specimen size and shape of the notch cannot be correlated with the results of other tests, including dynamic tension tests. The correlation of the notch impact value of a material with design is nonexistent; the correlation with performance in service is vague. There appears to be a certain empirical relation between the notch impact values of different materials and their tendency to fracture under dynamically applied triaxial stress. However, the impact values provide only a comparative rating; they must not be considered as interpretable information.

The principal purpose of the conventional notched-bar impact testing is to detect *velocity-sensitive* or *temperature-sensitive* behavior, that is, abrupt change from ductile to brittle fracture, as expressed by a sharp drop of the energy absorbed to fracture at a certain critical transition velocity or temperature. Since the behavior of the specimen also reflects the influence of the triaxiality of the state of stress at the root of the notch, which itself affects the type of fracture by determining the ratio between volumetric and distortional strain energy during the deformation and fracture process, the test does not provide a direct indication of either the effect of temperature or of velocity. The recorded function of impact values versus temperature or impact velocity express therefore the effect of the strain rate or the temperature of the test, combined with the effect of the geometrical characteristics of the specimen on deformation, on strength, and on the rate of crack propagation. The recorded transition temperatures or transition velocities can therefore be interpreted only as rating values of comparative behavior of different materials under *identical* testing conditions. Since the testing conditions of a notched-bar test cannot be quantitatively defined, transition temperatures or velocities obtained on notched specimens of different shape and size cannot be compared with each other, nor can they be quantitatively related to the performance of the material in service.

## 90. Fatigue Tests

The interpretation of the significance of results of conventional tests of specimens subjected to repeated load cycles is made difficult by the following facts:

1. That progressive damage is a process essentially determined by happenings on the submicroscopic scale, the cumulative effects of which become visible on a phenomenological scale at such an advanced stage of the fatigue test that most of the damage has already been done.
2. That the results of fatigue tests show a much wider scatter than the results of any other mechanical tests.
3. That the effects of the size and the shape of the specimens, of their surface condition, and of their environment (corrosion) are especially pronounced.

Moreover, the test results very much depend on the testing procedure and apparatus. The conventional testing conditions are hardly representative of the conditions of service, and the usual methods and techniques of representation of test results are inadequate. With regard to the latter shortcomings, however, fatigue tests are not different from other types of mechanical tests.

In the testing under repeated load cycles specimens are usually subject to either bending or axial loading or torsion; the cycle imposed may be one of full stress reversal between equal stress limits of unequal sign, or of stress fluctuation between unequal stress limits of equal or unequal sign. Every arbitrary cycle of fluctuating stress, however, can be split into a steady (mean) stress and a superimposed stress reversal, so that full stress reversal represents a condition of zero mean stress (Fig. 90.1). This splitting off of the mean stress is important with regard to the interpretation of the effects of inelastic deformation in fatigue tests, since it is the steady (mean) stress rather than the rapidly fluctuating cyclic stress that determines the intensity of plastic redistribution of stresses, or the effect of creep accompanying the fatigue process.

Of the different types of fatigue tests, bending tests are the easiest to perform and the most difficult to interpret, whereas the axial load tests are relatively easy to interpret but much more difficult to perform than bending tests. The principal difficulties with regard to the interpretation of fatigue test results

are the influences of surface conditions of the specimen and of environment, as well as the effects of inelastic deformation. These influences are most pronounced in bending tests and in torsion tests. In spite of this fact the simplicity of the bending fatigue test, particularly in the form of the rotating beam or cantilever test, has made it the standard fatigue test.

The influence of surface conditions and of environment are necessarily the more pronounced, the more important the contribution of the surface layers of the specimen to its fatigue performance. It is for this reason that the test results obtained in

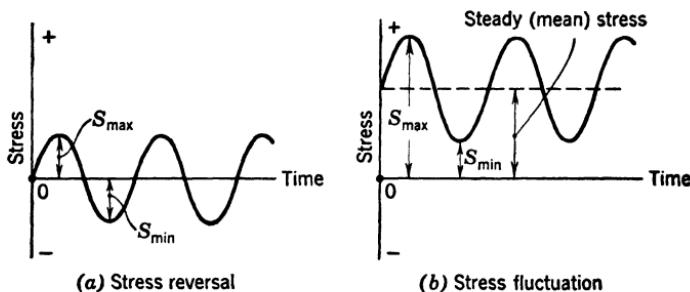


FIG. 90-1 Stress-time diagrams in fatigue tests.

bending and torsion tests of plain specimens, in which the maximum stresses are within the surface layers, depend very markedly on surface conditions and environment. Because of this dependence it is actually not clear to what extent bending and torsion fatigue tests represent tests of the surface conditions rather than of the fatigue performance of the material as such. If thin-walled specimens are used in those tests instead of plain ones, the effect of the stress gradient is largely eliminated; at the same time, however, the surface effects are intensified by the introduction of a second, inner surface, the condition of which, moreover, can be less well controlled than that of the outer surface, and is therefore less well reproducible. In addition to this difficulty, the anisotropy of specimens prepared from thin walled tubes is usually very pronounced, and cannot be eliminated.

Because of the uniformity of the stress distribution over the cross section of the specimen, the surface and environment effects are considerably reduced in axial load tests. However, such tests require a rather complex apparatus, and are compara-

tively difficult to perform because of the difficulty of producing adequate conditions of axiality of the loading and of controlling the operating frequency of the machine.

The significance of the influence of the inelastic deformation depends on the loading characteristics of the fatigue testing machine. If, as in the nonrotating bending machines and in most of the axial-load machines, the load cycle is produced with the aid of an oscillating spring which controls the deformation or the strain amplitude imposed on the specimen, the characteristic of the machine is similar to that of a *hard* tension machine used in static tests, if the rigidity of the spring is high in comparison to the rigidity of the specimen. An inelastic deformation of the specimen occurring in the course of the fatigue test compensates part of the elastic deformation of the spring, and thus reduces the amplitude of the imposed stress cycle. This "constant-strain" type of machine responds therefore to inelastic deformation of the specimen by a drop of the load, which is the more pronounced, the harder the spring controlling the load. In order to keep the stress amplitude constant during the test in a constant-strain machine, the deformation of the loading spring must be continually adjusted.

If the amplitude of the stress cycle is kept constant by application of a constant load, as in a rotating-beam machine, or practically constant by use of a very soft loading spring, as in a hydraulic pulsator, the characteristic of the machine is similar to that of a *soft* tension machine. This type of machine does not respond to inelastic deformation of the specimen by a change of stress.

In mechanical-oscillator types of fatigue machines the load cycle is produced by the centrifugal force resulting from an unbalanced rotating mass; the load amplitude may be controlled by either the eccentricity of the unbalanced mass or its speed of rotation, if the machine operates at a frequency different from the natural frequency of the specimen, or by the variation of the natural frequency of the vibrating system, if the machine operates at this frequency. Oscillator machines operating below the natural frequency of the specimen are essentially of the constant-stress type, since the force of the oscillating mass is controlled by the elastic resistance of the specimen and is therefore not affected by its inelastic deformation. The specimen is

subject to a practically constant stress amplitude and responds by a gradually changing deformation. Oscillator machines operating at a frequency substantially above the natural frequency of the specimen are essentially of the hard constant-strain type, since the amplitude of the load cycle is determined by the unbalanced mass of the oscillator; for constant mass the amplitude of the deformation cycle imposed on the specimen remains constant, the inelastically deforming specimen being thus subject to constant-strain cycles to which it responds by a varying stress. In order that oscillator-type machines be able to operate at the natural frequency of the specimen, the response of the specimen must be very nearly elastic: the permissible damping which can be compensated by the input energy is small, and, although these machines are essentially constant-deformation machines, the differentiation between constant strain and constant stress is immaterial under the conditions of near elasticity. These resonance machines, which may be incited mechanically, electromagnetically, or pneumatically, are usually shut off when in the course of the test the frequency of the specimen drops sharply below its initial natural frequency. However, it is not easy to decide whether the damaging effect producing such reduction of the natural frequency is the result of widespread inelastic deformation or of the development of a fatigue crack. In any case the definition of fatigue failure observed on a resonance machine must be different from that of fatigue failure observed on any other type of machine, which is defined by visible cracking or full separation of the specimen. Because of this fact, as well as because of the influence, through the inelasticity of the specimen, of the testing apparatus and procedure on the test results, the comparison of results of fatigue test obtained on different machines will frequently be misleading and should, therefore, be avoided.

The effect of the inelastic behavior of the specimen also accounts for the influence of the frequency of the imposed load cycles on the test results. When the duration of the individual load cycle is of an order of magnitude comparable to the order of magnitude of the relaxation times of the material, this influence is considerable both in viscoelastic materials and in metals. Under such conditions a combination of fatigue and creep effects is responsible for the progressive damage produced by the applied

load cycles. Because of this combination such fatigue-test results will be representative only if the frequency of the test is the same as the frequency of the load cycles under service conditions. Otherwise, the number of load cycles imposed on a specimen in the course of a test of a duration equal to the expected life of the considered structural or machine part will differ considerably from the number of cycles this part is expected to sustain in service. No correlation between test and performance under such conditions can be expected. This consideration is of particular importance with regard to machine parts operating at very high temperatures and vibrating at frequencies considerably above the operating frequencies of conventional fatigue-testing machines, such as turbine blades.

Although it is usually assumed that the fatigue performance of metals at room temperature is not affected by the frequency of the test, at least within the practical frequency range of fatigue machines, a certain frequency effect will necessarily be observable at relatively high frequencies. This is due to the fact that, since at constant frequency the average strain rate is proportional to the amplitude, it decreases with decreasing amplitude. At constant amplitude, on the other hand, it increases with increasing frequency. Since the yield limit is a function of the strain rate, it will change with changing frequency of the test; as far as the fatigue performance depends on the yield limit or is related to it, it will also change with frequency. This effect will become clearly observable, however, only if the difference between the compared frequencies is one of several orders of magnitude, because the yield point varies with the logarithm of the strain rate.

Because of the influence of the strain rate on fatigue performance the form of the load cycle should be expected to affect this performance. The load is usually not applied or removed at a constant rate, but at a rate varying periodically between a maximum at the mean stress and zero at the maximum and minimum stresses. No information is available concerning the effect of the shape of the load-time or stress-time diagram on the fatigue performance of the specimen.

The results of fatigue tests are usually presented in the form of graphs indicating the level  $S$  of the repeated stress cycle as a function of the number  $N$  of cycles sustained prior to failure,

the so-called  $S-N$  diagrams. These diagrams are, in general, determined by a rather arbitrary process of curve fitting through a relatively small number of points representing results of individual fatigue tests performed at various stress levels. In the presentation of  $S-N$  curves it is frequently not even stated whether these curves have been obtained as estimated "lines of best fit" or whether, following a procedure proposed by several investigators, they are curves drawn in such a manner as to remain below any of the individual test results. Evidently either of these procedures will lead to radically different  $S-N$  diagrams of different significance.

Since within the usual testing range any of the  $S-N$  curves, in general, can be approximated by a simple power law, it is transformed into a straight line in double-logarithmic scale. This scale is therefore most frequently used in the practical representation of fatigue results, the straight line relation being of some advantage in the drawing of the diagram through the test points. The representation of fatigue-test results in semilogarithmic scale will result in a straight-line relation only for brittle materials or for ranges of nearly brittle behavior (see Art. 60). Within ranges of large-scale inelastic deformation, however, even the double-logarithmic straight-line relation becomes unreliable, unless the tests have been performed in constant-stress machines, since the computation of the stress value from the imposed strain becomes the less reliable, the larger the inelastic strain component, and the more difficult therefore its compensation by adjustment of the loading spring. Fatigue-test results of metal specimen as represented by  $S-N$  diagrams are therefore the less reliable, the nearer the test stress approaches or the more it exceeds the yield limit of the material, true or designated. Hence, the softer the metal under the testing conditions, the more inadequate are the hard constant strain machines, and the more important the use of constant stress or soft constant-strain machines. This consideration is of importance with regard to fatigue tests on metals at elevated temperatures. In differentiating between a hard and a soft machine for this purpose, it should be borne in mind that the hardness or the softness of the constant-strain machine depends on the rigidity of the spring relative to the rigidity of the specimen. Hence, a machine that is soft at room temperature may be hard at the elevated tempera-

ture of the test, at which the rigidity of the specimen is reduced by creep.

The conventional  $S-N$  diagrams are intended to represent the average performance ("line of best fit") or the minimum performance (lowest line) in fatigue of the tested specimens, assumed to be representative of the material. The exponent of the power law in the natural scale representation, or the (negative) slope of the straight-line relation in double-logarithmic representation is a measure of the rate of progressive damage, since it indicates the reduction, with  $\log N$ , of the logarithm of the stress. It is therefore affected both by the severity of the stressing, as indicated by the large-scale stress concentrations (notches) and surface effects (corrosion pits), and by the resistance to progressive damage of the material, as indicated by the severity of the textural stresses. Thus, the slope of an  $S-N$  diagram in double-logarithmic representation is increased if either the stress concentrations of the external stress field or the inhomogeneities of the field of textural stresses are intensified; it is reduced as a result of all effects that reduce the inhomogeneities of the external stress field, such as plastic relief of peak stresses, or of the field of textural stresses, such as recovery after work hardening.

The conventional representation of the results of fatigue tests by a single  $S-N$  diagram is completely inadequate, since such representation neglects a particularly characteristic aspect of all fatigue data: their scatter. This scatter is much wider than the unavoidable scatter of all test results and observations, because the fatigue resistance of a material depends much more than any other mechanical property on the individual response to the imposed conditions of each element that makes up the specimen. Moreover, although in all other mechanical tests the testing conditions are imposed just once, they are periodically repeated in the fatigue test; therefore, the unavoidable statistical variation is also introduced into the testing conditions. It is this combination of the statistical effects in the applied load and in the response of the material that produces the wide scatter characteristic for results of fatigue tests. Extreme care in the preparation of specimens can be only partially effective in reducing this scatter, since only the surface conditions of the specimen are thus controlled. However, the more significant the contribution of the surface to the observed fatigue performance (for instance in

bending or torsion), the more effective the control of the surface conditions of the specimens in reducing the scatter of test results.

A considerable reduction of the scatter of fatigue-test data is achieved by plastic deformation, which produces an averaging effect with regard to both the textural stresses and the inhomogeneities of the imposed stress field. Hence, the scatter of the results of fatigue tests of metals will be the narrower, the more extensive the plastic deformation that precedes fatigue failure. The higher the intensity of the applied stress cycle, the narrower

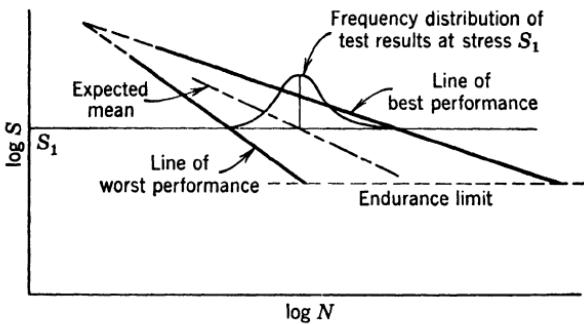


FIG. 90-2 Rational representation of results of fatigue tests.

therefore the scatter of the test results. These results should therefore not be presented in the form of a single  $S-N$  diagram, but by two converging straight lines in double-logarithmic scale, enclosing all results which might be expected to fall, with a designated probability, within the area between them; the steeper line indicates the limit of worst performance, the flatter line that of best performance (Fig. 90-2). However, the point of intersection of these two lines, which is usually located at a finite number of stress cycles, cannot be considered to be particularly significant, since the determination of the actual stress values at this level is usually not very reliable as a result of the large component of inelastic deformation.

If the number of test results is insufficient for a fairly correct evaluation of the position of the limiting  $S-N$  diagrams by statistical interpretation, the usual procedure is to represent the results by a single  $S-N$  diagram best fitting the test points. This procedure is highly unreliable, since the slope of this diagram depends more strongly on chance than the individual test points,

because of the large number of combinations of test points of equal probability of occurrence through which  $S$ - $N$  diagrams of different slopes, but equal probability, may be drawn. This fact is easily proved by the comparison of any number of  $S$ - $N$  diagrams for the same material, obtained from similar series of tests of relatively small numbers of "identical" specimens<sup>90-1</sup> (Fig. 90-3).

A partial remedy of this difficulty associated with the use, in conventional fatigue testing, of a relatively small number of specimens distributed over the whole stress range, can be obtained by performing the tests at only two stress levels, sufficiently distant to enclose most of the relevant testing range, so that the average of  $\log N$  at each stress level can be determined with the increased reliability provided by a larger number of test points at each level. If inelastic deformation can be assumed to reduce the scatter range at the higher stress level, the number of test results required at this level may be smaller than that at the lower stress level at which the scatter is necessarily wider. Thus, the relatively best results of a small number of fatigue tests will be obtained by testing about one quarter to one third of the total number of specimens at the highest stress level that can be reliably imposed by the machine, and the remaining specimens at a stress level low enough to produce larger scatter, but not too low to produce failure within reasonably short testing periods, the assumption being that within the testing range a straight-line  $S$ - $N$  diagram is obtained in double-logarithmic representation. The test results should then be presented by a line passing through the best possible estimates of the mean of  $\log N$  at both stress levels.

The shape of the frequency distribution  $p(N)$  of the fatigue strength at any level  $S$  cannot be determined from a small number of tests, but can only be estimated on the basis of previous knowledge concerning the character of this distribution, derived from the results of a large number of fatigue tests at various constant-stress levels. The small number of investigations of this type<sup>90-2</sup> have shown a characteristically skew frequency distribution  $p(N)$ , with the modal value substantially below the mean; however, this skewness is considerably reduced if  $p(\log N)$  is plotted instead of  $p(N)$ , the frequency distribution  $p(\log N)$  approaching the normal (Gaussian) distribution func-

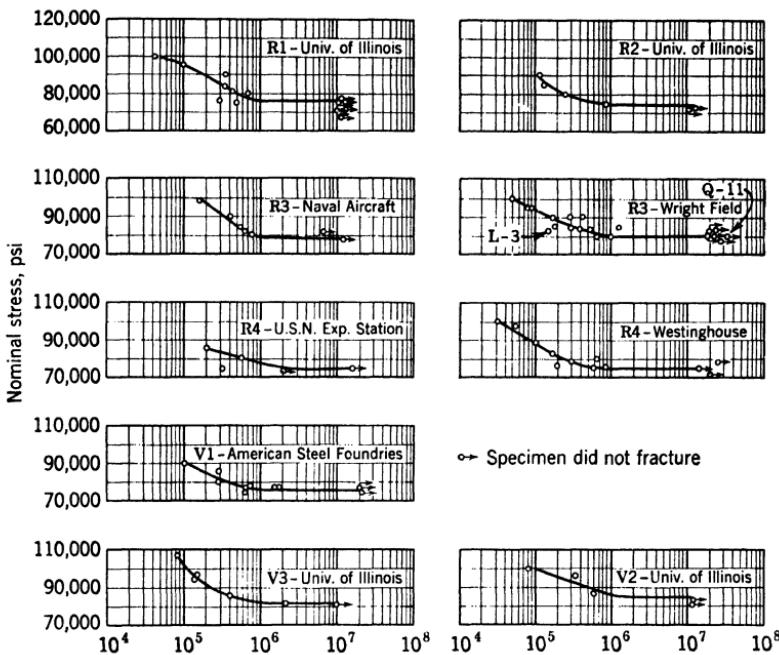


FIG. 90-3a Semilog plot of  $S$ - $N$  diagrams from data reported by ASTM Research Committee on Fatigue of Metals.<sup>90-1</sup>

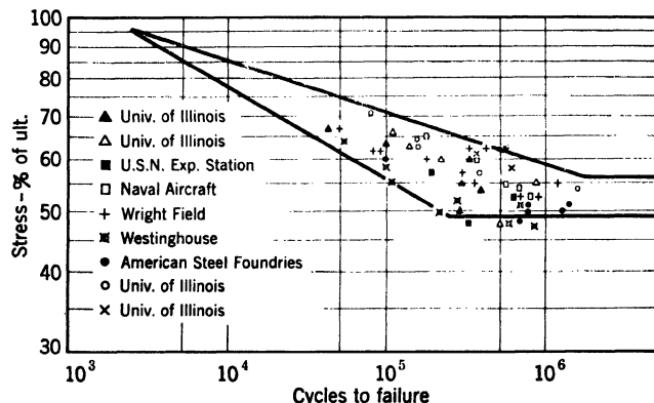


FIG. 90-3b Log-log replotting of data of Fig. 90-3a and estimated lines of best and worst performance (after Almen<sup>90-1</sup>).

tion fairly well. By representing the cumulative distribution function  $P(N) = \int_0^N p(N)$ , instead of the frequency distribution, the similarity between the fatigue data at a certain level of imposed stress and the so-called "mortality data" as expressed by life-expectancy charts of various machine parts, or of other products in which "life" is an important factor, such as light bulbs and radio tubes, becomes obvious<sup>90-3</sup> (Fig. 90·4). Fatigue

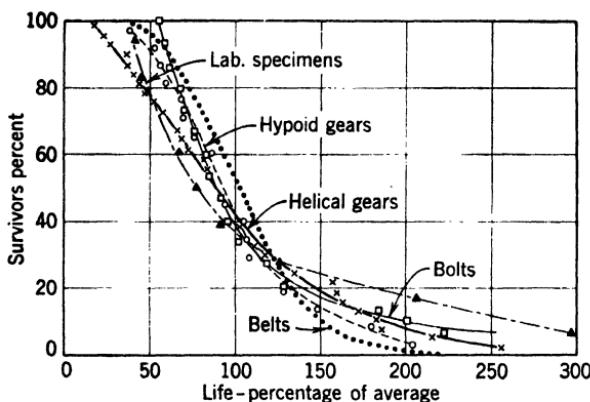


FIG. 90·4 Life expectancy (mortality) curves of various machine parts and of laboratory fatigue specimens (after Almen<sup>90-3</sup>).

data, therefore, should be considered in the same light as "mortality data" and interpreted as such. The general realization of this fact by designers will do away with the tacit assumption that fatigue strength is a characteristic that can be as accurately determined as the yield stress or the "ultimate tensile strength," and can therefore be related to either of these values by a simple multiplier.

For metals showing a definite endurance limit the  $S-N$  diagram, which is defined in terms of failure occurring after a finite number of load cycles, loses its significance at this limit, which is defined by the nonoccurrence of failure. This difference in the significance of the  $S-N$  diagram and the endurance limit makes the conventional plotting of an extension of the  $S-N$  diagram as a line parallel to the  $N$  axis within the range of indefinite endurance rather meaningless; the so-called "knee" in the double-logarithmic  $S-N$  diagram, which assumedly divides the finite life

range from the endurance range, is in fact the limiting point of the validity of the  $S$ - $N$  diagram, since this diagram expresses a significant relation only within the range of stresses under the repeated application of which the specimen actually breaks or otherwise fails. This relation is derived from the results of tests of a "population" of specimens *breaking* at various combinations of stress level and number of repetitions. The endurance limit, on the other hand, is the highest stress level at which the specimen is not expected to fail even after an infinite number of stress repetitions; therefore, it can be *estimated* only from an extrapolation to zero or to a designated measure of probability of the observed *percentage* of specimens failing at various stress levels. Still better, it can be estimated by a procedure known as "sensitivity testing"<sup>90, 1</sup> in which the specimens are subjected to stress intensities ranging from a level at which practically no failures occur to a level at which almost all specimens fail, the stress for any given specimen being determined by whether the preceding specimen failed or did not fail. The usual procedure of presenting the endurance limit as a line parallel to the  $N$  axis, drawn below any point representing a test at which a specimen has "run out," that is, at which it has not failed after a practically infinite number of load repetitions, cannot be expected to lead to a correct estimate of the true endurance limit of the material.

If no endurance limit exists because of the influence, under the conditions of the test, of time-dependent processes, such as creep, diffusion (recovery and precipitation), recrystallization, or corrosion, the  $S$ - $N$  diagrams do not attain their lower limit of validity within the range of finite values of  $N$ ; however, because of this so extended range of validity of the diagram, its whole range usually cannot be represented by a single power law, and thus by a straight line in double-logarithmic representation, but only by several power laws valid within certain parts of the range. Hence a curved  $S$ - $N$  diagram is obtained even in double-logarithmic representation, asymptotically approaching the intrinsic endurance limit determined by the slip resistance of the largest crystal size present in the polycrystalline aggregate (see Art. 62).

The sharp differentiation for metals that have an endurance limit, between the finite life range of fatigue performance in which the  $S$ - $N$  diagram is valid and the range of indefinite

endurance, is of considerable practical importance. The influence of specimen shape and size, as well as the influence of temperature and of history (heat treatment, previous cold work, surface treatment) will necessarily be different with regard to the location and slope of the  $S-N$  diagram, and with regard to the endurance limit, since in the first case the relevant effects are those on crystal fragmentation, crack initiation, and propagation, and in the second those on the process of blocking of slip and crystal fragmentation. Although in both cases the most significant part of the influence of the various effects is on the inelastic deformation preceding and accompanying fatigue, the fatigue performance in the first case depends on the *progress* of such deformation, and in the second on its *blocking* at an early stage.

Thus, for instance, the effects of size and shape of fatigue specimens must necessarily influence the slope of the  $S-N$  diagram more than it influences the endurance limit. This conclusion is borne out by comparative torsion fatigue tests of round solid and hollow steel specimens of equal diameter, which showed that the slope of the  $S-N$  diagram of the hollow specimen was much steeper than that of the solid specimen, because of the smaller cross-sectional area to be destroyed, whereas the endurance limit of both types of specimens was identical.<sup>90-5</sup> This is only to be expected, since at the maximum fiber stress, at which the cold-working effects can still be rapidly blocked and the response of the specimen become elastic, which is the stress at the endurance limit, no reason exists for a different performance of a plain and a hollow specimen, as only the response of the surface layer is significant. On the other hand, prestraining in torsion of such specimen might possibly affect the endurance limit more than the  $S-N$  diagram, since the resulting raising of the yield point will substantially raise the endurance limit, while the life of the specimen at a definite stress cycle may be much less affected.

Since the inelastic behavior of the material subject to repeated stresses below the yield limit is not related to the inelastic deformation associated with the results of other tests, except the damping test (see Art. 62), the results of repeated load tests, in general, cannot be correlated with the results of other mechanical tests. Moreover, as long as the inelastic deformation and the

damage during the test can be produced only by load repetition, there appears to be no other possibility of determining the fracture stress for a certain number of load cycles than subjecting the material to this number of load cycles. Only when, as in high-temperature testing, the progressive damage is produced by both load repetition (cycle effect) and creep (time effect) does it become possible to find a certain correlation between the results of repeated-load tests and creep tests. This is due to the fact that the same localized inelastic deformation that is produced by a repeated load might be produced by a constant load of different intensity sustained for a certain period.

A shortening of repeated-load tests at temperatures at which creep effects are negligible can thus be achieved only by increasing the frequency of the load cycles. However, in doing this the effect on the inelastic deformation of increasing the strain rate must be considered.

The significance of the results of fatigue tests in terms of structural performance would be obvious, if the loading conditions of the test were to reproduce those in service, as in this case a direct correlation would exist between test and performance. Such conditions are, however, the exception rather than the rule; so far no reliable and practical basis has been developed for the interpretation of the significance of the results of conventional fatigue tests (load cycles of constant amplitude) in terms of actual performance under load cycles of widely varying amplitudes and different sequence of application interrupted by periods of rest.

### References

- 86·1 Symposium on the Significance of the Hardness Test, *Proc. ASTM* **43** (1943) 803.
- 86·2 *Ibid.*, 804.
- 87·1 C. W. MACGREGOR, *Proc. ASTM* **40** (1940) 515.
- 87·2 W. KUNTZE and G. SACHS, *Z. Ver. deut. Ing.* **72** (1928) 1011.
- 87·3 G. SACHS and G. FIEK, *Der Zugversuch*, Leipzig, Akad. Verlagsges. (1926) 40.  
H. O'NEILL, *Proc. Inst. Mech. Engrs. London* **151** (1944) 116.
- 88·1 A. SAUVEUR, *Proc. ASTM* **38** (1938) 10.
- 88·2 F. B. SEELY, *Proc. ASTM* **40** (1940) 541, 547.
- 89·1 Symposium on Impact Testing, *Proc. ASTM* **38** (1938) 105.
- 90·1 J. O. ALMEN, *SAE Trans.* **51** (1943) 265.

90·2 H. MUELLER-STOCK, *Mitt. Kohle u. Eisenforschung* **2** (1938) 83.

A. M. FREUDENTHAL, C. S. YEN, and G. M. SINCLAIR, *8th Progress Rept. on Fatigue*, Office of Naval Research & Univ. Illinois Eng. Expt. Sta., Urbana (1948).

90·3 J. O. ALMEN, *op. cit.*, 262.

90·4 B. EPSTEIN, *ASTM Bull.* **158** (1949) 62.  
J. T. RANSOM and R. F. MEHL, *Metals Trans.* **185** (1949) 364.

90·5 H. J. GOUGH, *The Fatigue of Metals*, Scott, Greenwood & Son, London (1924) 83.

## AUTHOR INDEX

Alfrey, T., 108, 164, 230, 248, 356  
 Almen, O. J., 569, 570, 573, 574  
 Andrade, N. da C., 164, 316, 317, 325  
 Arrhenius, S., 113

Barrett, C. S., 397  
 Batdorf, S. B., 505  
 Becker, R., 340, 356  
 Beliaev, N. M., 437, 439, 457  
 Bernhard, R. K., 355  
 Betty, B. B., 325  
 Bijlaard, P. P., 504, 505  
 Bingham, E. C., 221, 403, 412  
 Biot, M. A., 247  
 Birch, F., 248  
 Bleich, F., 493, 495, 505  
 Boas, W., 108, 164  
 Bocker, R., 505  
 Bohr, N., 35, 36  
 Boltzmann, L., 112, 113, 135, 191,  
     338, 339, 341, 355  
 Born, M., 26, 60, 164, 213  
 Bragg, L., 70, 71, 96, 107  
 Brandtzaeg, A., 505  
 Brick, R. M., 164  
 Bridgman, P. W., 164, 248, 394,  
     397  
 Brillouin, L., 187  
 Buckingham, E., 413  
 Burgers, J. M., 237

Cherniak, G. S., 325, 526  
 Chwalla, E., 505  
 Clark, S. D., 356  
 Coffin, L. F., 325, 526  
 Cottrell, A. H., 164  
 Coulomb, C. A., 260, 279  
 Cramér, H., 361, 397

Daniels, H. E., 397  
 Datwyler, G., 356

Davis, E. A., 279, 280, 463  
 De Broglie, M., 44, 45  
 Debye, P., 85, 164  
 Dehlinger, U., 95, 108  
 Dirac, P. A. M., 44  
 Dokos, S. J., 463  
 Dolan, T. J., 397  
 Dollins, C. W., 317, 325  
 Donell, L. H., 440, 451  
 Dorn, J. E., 213  
 Dunbar, L. W., 213  
 Dushman, S., 213

Eichinger, A., 258, 279  
 Einstein, A., 400, 413  
 Eisenschitz, R., 213  
 Elam, C. F., 145, 164, 299, 300,  
     304  
 Engesser, F., 500, 501, 505  
 Epstein, B., 397, 574  
 Eyring, H., 108, 164, 248

Fick, G., 573  
 Fisher, J. C., 164, 213  
 Fisher, R. A., 361, 397  
 Fraenkel, S. J., 280  
 Frank, N. H., 213  
 Fredrickson, J. W., 164  
 Frenkel, J. I., 397  
 Freudenthal, A. M., 26, 304, 397,  
     451, 505, 574  
 Freundlich, H., 401, 413  
 Fritzsche, J., 450, 505  
 Fromm, H., 269, 280  
 Fuerth, R., 397  
 Fullman, R. L., 213

Gauss, L., 48  
 Geil, G. W., 397  
 Geiringer, H., 454, 463  
 Geisler, A. H., 304

Gibbs, J. W., 112, 135, 193  
 Glasstone, S., 108, 164  
 Goetz, A., 108  
 Goldberg, A., 213  
 Goss, N. P., 108  
 Gough, H. J., 164, 574  
 Grammel, R., 335, 355  
 Griffith, A., 358, 359, 360, 364, 397  
 Griggs, E., 396  
 Grimm, H., 107  
 Gruening, M., 487, 505  
 Guinier, A., 108  
 Guth, E., 26  
 Haar, A., 267, 280, 430  
 Haase, O., 164  
 Haegg, G., 60  
 Halsey, G., 248  
 Hamel, E., 248  
 Handelman, G. H., 280  
 Hanson, D., 164, 298, 304  
 Harrington, R. H., 60  
 Hedvall, J. A., 164  
 Heidenreich, R. D., 164  
 Heisenberg, W., 44, 48, 49  
 Hencky, H., 183, 187, 221, 222, 248, 255, 257, 267, 274, 279, 280, 430, 440, 451, 455, 463  
 Hill, R., 279  
 Hodge, P. G., 451  
 Hollomon, J. H., 108, 164, 213, 350, 356, 397  
 Hooke, R., 21, 222, 232  
 Hoppmann, W. H., 353, 356  
 Houwink, R., 75, 107, 247, 414  
 Huber, M. T., 255, 279, 389  
 Hudson, D. R., 60  
 Huthsteiner, H., 213  
 Inglis, C. E., 359, 397  
 Jackson, L. R., 273, 280  
 Jaffe, L. D., 108, 164  
 James, H. M., 26  
 Jeffrey, G. B., 26  
 Jenkin, F. C., 248  
 Jenkins, W. H., 397  
 Kantorova, T. A., 397  
 Kármán, Th. v., 267, 280, 351, 352, 356, 397, 430, 455, 458, 463, 485, 500, 501, 505  
 Karnop, R., 164  
 Ké, T. S., 164, 344, 345, 348, 356  
 Kehl, G. L., 213  
 Kelvin, Lord, 220, 248  
 Kirkwood, J. G., 108  
 Kochendoerfer, A., 164  
 Koehler, J. S., 96, 108  
 Kohlrausch, F., 248  
 Kohlrausch, R., 248, 338, 355  
 Kollbrunner, C. F., 450, 498  
 Kuntze, W., 573  
 Lacombe, P., 108  
 Laidler, K. J., 108, 164  
 Lankford, W. T., 280  
 Lederman, H., 248, 355  
 Lee, E. H., 279, 356  
 Lin, C. C., 280  
 Lode, W., 279  
 Lubahn, J. D., 463  
 Ludwik, P., 83, 108, 182, 187, 208, 213, 546, 553  
 Maerea, A. E., 451  
 Magnel, G., 526  
 Maier-Leibniz, H., 498, 505  
 Majors, H., 397  
 Malaval, M., 451  
 Masing, G., 108  
 Matossi, F., 108  
 McAdam, D. J., 391, 397  
 McGregor, C. W., 164, 213, 397, 543, 546, 573  
 McReynolds, A. W., 150, 164  
 Mebs, R. W., 397  
 Mehl, R. F., 574  
 Miller, R. F., 213  
 Millis, B. D., 397  
 Mindlin, R. D., 231, 248  
 Mises, R. v., 60, 255, 256, 262, 265, 279  
 Mohr, O., 181, 279, 393  
 Mooney, M., 248  
 Moore, H. F., 317, 325

Mott, N. F., 397  
Mueller-Stock, H., 574  
Muir, J., 164  
Murnaghan, F. D., 248

Nabarro, F. R. N., 108, 164  
Nádai, A., 140, 164, 258, 279, 325,  
    419, 439, 440, 445, 450, 451, 453,  
    455, 456, 459, 463, 465, 470, 471

Nix, F. C., 108  
Nye, J. F., 107, 304

O'Neill, H., 573  
Orowan, E., 95, 101, 108, 213, 455,  
    463

Oseen, C. W., 453, 463

Panov, D., 248  
Pauli, W., 37, 39  
Pipes, A. L., 60, 355  
Planek, M., 32, 33, 36, 114, 191  
Polanyi, M., 95, 108  
Popov, E. P., 325  
Poynting, J. H., 247  
Prager, W., 264, 265, 269, 271, 276,  
    278, 280, 450, 485, 496, 505  
Prandtl, L., 95, 108, 140, 164, 268,  
    418, 444, 445, 446, 447, 451, 453,  
    463

Putnam, W., 279

Quinney, H., 271, 280

Rabinowitseh, B., 213  
Rachinger, W. A., 108, 325  
Ransom, J. T., 574  
Read, T. A., 108  
Rehkuh, F., 248  
Reiner, M., 26, 213, 221, 248, 304,  
    409, 413, 414  
Reiner, R., 26  
Reuss, E., 268, 280  
Rice, F. O., 26, 60, 107  
Rivlin, R., 414  
Roš, M., 258, 279  
Roseoe, R., 164  
Rosenhain, W., 164

Ross, A. D., 519, 526  
Rutherford, E., 36

Sachs, G., 164, 451, 463, 573  
Sadowsky, M. A., 280  
St.-Venant, B. de, 221, 260, 271, 279

Sauveur, G., 573  
Sayre, M. F., 304  
Schaeffer, C., 108  
Schmid, E., 164  
Schroedinger, E., 44, 46  
Scott-Blair, G. W., 414  
Seely, F. B., 279, 426, 450, 573  
Seitz, F., 108, 164  
Seth, B. R., 248  
Shaffer, B. W., 356  
Shanley, F. R., 502, 505  
Shepler, P. C., 325, 526  
Shockley, W., 69, 107, 108, 164  
Siegfried, W., 213  
Sinclair, G. M., 574  
Sinitzki, A. K., 437, 438, 439, 451  
Slater, J. C., 213  
Smekal, A., 95, 108, 164  
Smith, J. O., 426, 450  
Smith, K. F., 280  
Smith, R. B., 397, 510  
Smith, S. I., 304  
Smith, V. G., 213  
Smoluchowski, M. v., 400, 413  
Sokolnikoff, I. S., 187, 450  
Stead, O., 547, 548  
Sternberg, E., 247  
Stuessi, F., 498  
Swainger, K. H., 304, 471  
Swift, W., 187

Taylor, G. I., 95, 108, 164, 271, 275,  
    280, 304, 351, 356  
Teller, E., 26, 60, 107  
Terzaghi, K., 414  
Thomson, J. J., 247  
Thum, A., 450  
Tietz, E., 213  
Timoshenko, S., 187, 451  
Tippett, L. H. C., 361, 397  
Tolman, R. C., 108, 163  
Tracy, D. P., 463

**Author Index**

Tresca, H., 260, 279  
Trenting, R. G., 164  
Tupper, S. J., 279  
Turnbull, D., 213  
  
Voigt, W., 338, 355  
  
Walther, H., 355  
Weber, W., 338, 355  
Wegel, R., 355  
Weibull, W., 397  
Weiser, H. B., 413  
Weissenberg, K., 194, 213  
Wheeler, M. A., 164, 298, 304  
  
White, G. N., 451  
Wiechert, E., 245, 248, 338, 339, 355  
Wolff, H., 107  
Wood, W. A., 100, 108, 304, 325  
Woolley, R. L., 304  
Wunderlich, F., 450  
  
Yen, C. S., 574  
  
Zachariasen, W. H., 108  
Zener, C., 26, 213, 338, 339, 345, 348,  
          350, 356, 395, 397  
Zwickly, F., 96, 108  
Zwikker, C., 60, 108

## SUBJECT INDEX

Activation, 80, 110, 122*ff*, 145*ff*  
Activation energy, 84, 110, 117, 122*ff*,  
285  
determination of, 346*ff*  
effect of, on fatigue, 383  
on fracture, 383  
of diffusion processes, 120  
of iron, 344  
Activation rate, 136  
Adiabatic effects, in metal cutting,  
354*ff*  
in penetration of armor plates, 350  
on stress-strain diagrams, 217,  
349*ff*, 354, 557  
After-effects, 202, 205, 222*ff*  
superposition of, 307, 319  
Age hardening, 120*ff*, 301*ff*  
Alkali metals, 41, 54  
Allotropic modifications, 57, 92*ff*,  
135  
Alloys, 57*ff*  
Aluminum, deformation of, 150, 298  
Amorphous substances, 103, 125*ff*  
Analogy, creep-work hardening, 322  
electrical, 329*ff*  
membrane, 418*ff*  
sand-heap, 419  
viscoelastic, 321*ff*, 509, 516  
Anelastic characteristics, 345*ff*  
Anions, 43, 63  
Anisotropy, formation of, 140*ff*, 287*ff*  
of plastic deformation, 271*ff*  
of work-hardened metal, 380*ff*  
strain, 154  
Antimony, 65, 72, 90  
Argon, 41, 53  
Armor plates, penetration of, 350  
Asbestos, 71  
Asphalt, 410*ff*  
Atom, model of, 35*ff*, 46, 51  
quantum-statistical concept of, 51  
Atomic number, 27, 54*ff*  
Atomic radius, 52, 54*ff*  
Atomic weight, 61  
Austenite, 57  
Avogadro's number, 61  
Bauschinger effect, 275*ff*, 286, 381  
influence on fracture of, 392  
Beat frequencies, 31, 46  
Bifurcation of equilibrium, 503  
Binding energy, 75  
mass equivalence of, 7  
Bingham body, 269, 403  
Blocking of slip planes, 120*ff*, 130,  
149*ff*, 283*ff*, 298, 379*ff*, 428, 500  
Block-structure, 97*ff*  
Boltzmann-Maxwell distribution,  
111*ff*, 135  
Boltzmann-Planck relation, 191  
Bonds, types of, 61*ff*  
Brinell hardness, 536*ff*  
Buckling, plastic, 500*ff*  
viscoelastic, 518*ff*  
Bulk modulus, 216  
adiabatic effect on, 217  
effect of work hardening on, 281  
Canonical distribution, 112  
Carbon, 66  
diffusion of, 56  
Carburizing, 92  
Carrying capacity, 481  
of curved beams, 426  
of redundant plastic structures,  
481*ff*, 491*ff*, 497*ff*  
of trusses, 487*ff*  
Cations, 43, 63  
Cement, 105*ff*  
viscosity of, 523  
Ceramic materials, 105*ff*  
Characteristic temperature, 115*ff*

Circular hole, glide-line field of, 454  
     plastic solution of, 439

Clay, 412*ff*

Cleavage fracture, 163, 374

Closed electron shells, 41*ff*, 51*ff*, 64*ff*,  
     73

Cohesive strength, 75, 357*ff*  
     in tension test, 546, 551  
     statistical theories of, 360*ff*

Compressibility, 41, 54, 154*ff*

Conservation, of energy, 34, 189  
     of momentum, 34

Consistency diagrams, 402*ff*

Contained plastic deformation, 273*ff*,  
     276, 278, 481*ff*

Coordinates for definition of strain,  
     178

Coordination lattice, 63, 86*ff*

Copper, 69

Correspondence principle, 34

Coulomb attraction, 36, 63

Covalent bonds, 67, 71*ff*, 90

Cracking, delayed, 394*ff*

Cracks, atomic and submicroscopic,  
     159*ff*, 358*ff*, 373, 381  
     in creep processes, 142, 316*ff*

Crazing, 142

Creep, general theory of, 148*ff*  
     of concrete, 525  
     of lead, 313, 316*ff*  
     of single crystals, 314  
     periodicity of, 315  
     temperature dependence of, 149*ff*,  
         211, 315  
     three-axial, 311*ff*, 512  
     third stage of, 316*ff*  
     transient, 239*ff*, 307*ff*, 315*ff*, 509

Creep design, for short life, 509  
     vibratory stresses in, 509*ff*

Creep fracture, 382*ff*

Creep function, 319*ff*, 339

Creep recovery, 309

Creep strength, temperature dependence of, 371*ff*  
     time dependence of, 369*ff*

Crystal boundaries, formation of,  
     101  
     viscosity of, 101, 127*ff*, 344*ff*

Crystal classes, 89

Crystal fragmentation, 140, 284*ff*,  
     291*ff*

Crystal fragments (crystallites), 100,  
     140, 284, 380

Crystallization, 103  
     of high polymers, 141*ff*

Crystals, types of, 63*ff*, 70, 73, 100

Crystal systems, 89

Cubic lattice, 90*ff*

Curved beams, carrying capacity of,  
     426

Cyaniding, 92

Cylindrical hole, plastic solution of,  
     439, 454

Damage, functional, 474*ff*, 506*ff*  
     structural, 474*ff*

Damping, effect on, of fatigue, 336*ff*,  
     572*ff*  
     of grain size, 347*ff*  
     of temperature, 337, 348  
     variation of, 332*ff*, 343*ff*

Damping tests, purpose of, 332, 334,  
     337*ff*, 572*ff*

Dashpot, 232

De Broglie equation, 45, 49

Debye function, 114*ff*

Debye temperature, 115*ff*, 136

Decay function, 245

Deflection theory of arches, 520

Deformation, contained plastic,  
     273*ff*, 276, 278, 481*ff*  
     delayed, 121  
     discontinuous, 149*ff*, 286, 298  
     structural mechanism of, 122*ff*  
     temperature sensitivity of, 126

Deformation machines, 530

Deformation stresses, 515*ff*

Density, 55  
     of work-hardened metal, 285

Depolymerization, 142, 309

Deviator, of strain, 180  
     of stress, 174

Diamond, specific heat of, 115  
     structure of, 66

Diffusion, 56, 103

Diffusion, blocking of slip planes by, 120*ff*, 130, 149*ff*, 283*ff*, 298, 379*ff*, 428, 500  
coefficient of, 117*ff*  
grain boundary, 128, 286  
rate of, 117*ff*, 122  
strain aging, 120*ff*, 301*ff*  
types of, 118*ff*

Dilatancy, 220

Dipoles, 58

Discontinuity, of phenomena, general, 17  
of slip, 149*ff*  
of stress-strain diagrams, 286, 298  
of yielding, 286

Dislocations, 95, 124*ff*  
concentration of, 95, 97, 288  
hardening effect of, 144, 287*ff*  
soap-bubble model of, 96

Disorder, long-range, 85

Disorientation of crystal structure, 313

Divisibility of matter, 6*ff*

Ductility, measure of, 549*ff*

Elastic constants, 75, 216*ff*

Elasticity, nonlinear, 214, 218*ff*

Elastic limit, 22, 138

Elastic potential, 215*ff*

Electron-cloud density, 50*ff*, 58

Electronegative elements, 67

Electron microscope, 44

Electron orbits, 36*ff*, 52  
ellipticity of, 38  
energy level of, 37, 40  
filling of, 41*ff*

Electrons, excited state of, 39  
free, 68, 70, 73  
sharing of, 64*ff*, 73

Electron shells, closed, 41*ff*, 51*ff*, 64*ff*, 73  
incompletely filled, 69, 73

Electron spin, 39

Elliptical hole, stress concentration due to, 359

Endurance limit, 336*ff*, 383*ff*, 570*ff*

Energy, free, 192, 216

Energy, latent, 275, 285*ff*, 292, 335

Energy content of solid, 7, 82*ff*

Energy quantum, 32

Energy transformation, representation of, 194*ff*  
types of, 199

Entropy, 190*ff*, 208

Equations of state, existence of, 168, 208*ff*  
linear, 222

Equicohesive temperature, 212

Equilibrium, bifurcation of, 503  
conditions of, 170*ff*, 186

Equipartition theorem, 82*ff*

Equivalent elastic strain, 231

Exclusion principle, Pauli's, 37*ff*, 42*ff*, 64, 90

Extreme values, statistical theory of, 361*ff*

Fatigue data, scatter of, 567*ff*  
presentation of, 565*ff*  
significance of, 573

Fatigue strength, 382*ff*  
effect on, of cold work, 384, 572  
of creep, 385*ff*  
of frequency of loading, 387, 563*ff*  
statistical aspects of, 366*ff*  
under combined stresses, 391

Fatigue testing, 560*ff*

Ferromagnetism, 69

Finite strain, 219*ff*

Flocculation, 400

Flow, 9*ff*, 139  
theories of plastic, 263*ff*  
units of, 83, 86

Fluidity, 232, 403

Force machines, 530

Forces, interatomic, 73*ff*  
internal, 8

Foreign atoms, precipitation of, 101*ff*, 118, 314

Fracture, atomic concept of, 157*ff*  
conditions of, 157, 211*ff*, 390*ff*, 514*ff*  
delayed, 394*ff*

Fracture strength, brittle, 357*ff*

Fracture strength, effect, of anisotropy on, 392*ff*  
of hydrostatic stress on, 160*ff*, 393  
of inelastic deformation on, 158*ff*, 212, 381*ff*, 391*ff*  
of strain rate on, 212, 381  
of stress condition on, 391  
of temperature on, 381  
of time on, 377*ff*  
of volumetric strain on, 388*ff*  
of amorphous materials, 126, 387  
of bundles of threads, 367  
reproducibility of, 382

Fragmented crystal structure, 140, 284*ff*, 291*ff*, 313  
thermal instability of, 284, 291

Free electrons, 68, 70, 73

Free energy, 192, 216

Functional damage, 474*ff*, 506*ff*

Galton board, 47*ff*

Gaussian distribution, 48, 361*ff*

Glass structure, 104

Glide lamellas, 131*ff*, 143, 282

Glide laminas, 130*ff*

Glide lines, distance of, 132  
mathematical properties of, 250, 453  
propagation of, 251*ff*

Grain boundaries, cracking of, 317*ff*  
creep of, 312*ff*  
foreign atoms in, 101*ff*, 128, 286  
relaxation times of, 313  
viscosity of, 101, 344*ff*

Gram-atom, 61

Granular material, failure of, 394

Graphite, 65, 71*ff*

Group pattern, 16*ff*

Halogens, 41, 64

Hardening modulus, 535*ff*

Hardness, Brinell, 536*ff*  
definition of, 534*ff*  
rebound (scleroscope), 539*ff*

Heat, of sublimation, 64, 69  
specific, 82*ff*, 115

Helium, 7, 37, 40

Heterogels, 408

Heterogeneous materials, 106*ff*  
fracture of, 393

Hexagonal crystal structure, 65, 90  
slip delay in, 286

High elasticity, theories of, 218*ff*

High-temperature service, metals for, 508, 511

Homogeneity, statistical, 13

Homologous temperatures, 83*ff*

Hydrogen, 37, 50*ff*

Hyperbolic sine relation of inelasticity, 136, 139, 246

Hysteresis, 275*ff*, 326, 329

Impedance, 329*ff*, 342

Impenetrability of matter, 39, 56, 73

Indentation, 534*ff*

Inelasticity, thermal and athermal, 136*ff*

Inert gas, 41, 52, 73

Infrared radiation, 32, 80

Intercrystalline regions, viscosity of, 101, 127*ff*, 344*ff*

Interstitial alloys, 57

Interstitial diffusion, 119

Interstitial solid solution, 92

Invariants, 173*ff*, 186*ff*, 215*ff*, 270*ff*

Ionic bonds, 74

Ionic lattice, 86

Ionization, 70

Ions, 43, 68  
radii of, 52*ff*

Iron, allotropic modification of, 57, 93, 135

Isogels, 408

Isotropy, statistical, 13, 87, 103

Joule effect, 218

Kelvin body, deviator relations of, 223*ff*  
model of, 234

Kelvin elements, superposition of, 238, 307, 319

Krypton, 41

Laminar slip, 130*ff*, 282

Laminates, 107  
Lattice defects, 94*f*  
concentration of, 95, 97, 288  
Lattice vibrations, 76*f*  
Latent energy, 275, 285*f*, 292, 335  
Limiting resistance, 474  
Limiting size of crystal fragments, 100, 140, 284, 380  
Linear substances, classification of, 221*f*  
Liquid, definition of, 8, 196*f*  
undercooled, 85, 104*f*, 196  
Lithium, 40  
Load stresses, 515*f*  
Logarithmic decrement, 327  
Logarithmic strain, 183  
Lueder's lines, 250*f*

Machines, deformation and force, 530  
Macromolecules, 60*f*, 67, 73  
Martensitic transformation, 93, 135  
Mass defect, 7, 61  
Mass phenomena, definition of, 47  
Material body, 6*f*  
Maxwell body, beam equation of, 516  
deviator relations of, 223*f*  
Maxwell-Boltzmann distribution, 111*f*, 135  
Maxwell elements, nonlinear, 246*f*  
superposition of, 238*f*, 307  
Mechanical testing, purpose of, 527*f*  
Melt, solidification of, 98  
Melting process, 80*f*, 103, 368  
Membrane analogy in torsion, 418*f*  
Memory effects, 236*f*  
Mercury, 85  
Metal cutting, 354*f*  
Metallic bonds, 62, 68*f*  
Metals, alkali, 41, 54  
transition, 69  
Metal structure, soap-bubble model of, 70*f*  
Methacrylates, 141  
Mica, 71*f*  
Microcracks, 159*f*, 358*f*, 373, 381  
Miller indices, 89

Model elements, mechanical, 232, 235  
Models, multiple-element mechanical, 238, 242*f*, 246*f*  
Mohr's circles, 181, 261, 266  
enveloping lines of, 393*f*  
Mole, 61  
Molecular compounds, heat of sublimation of, 73  
Molecular crystals, 63, 73, 86  
Molecular suspension, 398*f*  
Molecular weight, 61  
Molecules, 58*f*, 64*f*, 72  
Mosaic structure, 96*f*

Necking in tension tests, 317, 543*f*  
Neon, 40  
Neutral change of state of stress, 264*f*, 504  
Nitriding, 92  
Nitrogen, 56  
Nonlinear elasticity, 214, 218*f*  
Nonlinear Maxwell element, 247  
Nonlinear stress distribution, 466*f*, 469*f*, 515, 524  
Nonlinear stress-strain diagram in design, 473, 476  
Notched-bar impact test, 558*f*  
Notch effect in tension test, 553*f*  
Nucleus, 37, 52

Octahedral shear strain, 181*f*, 258  
Octahedral shear stress, 176, 258  
Orbital motion, 36*f*  
Order, 24, 80, 84*f*  
Order-disorder, 85*f*, 102*f*  
Order-disorder spectrum, 127, 192  
Order-disorder transformation, 104  
Orthotropic plastic deformation, 272*f*  
Oscillators, 30*f*, 76, 114  
Overstress, resistance against, 549  
tensor of, 269

Paints, 412  
Paracrystalline formation in melts, 102  
Paraffin, 73

Particle size, 53  
 Pauli's exclusion principle, 37*ff*, 42*ff*,  
     64, 90  
 Penetration, of armor plate, 350  
     of plastic zone, 253*ff*, 425*ff*, 438  
 Periodic table of elements, 37*ff*  
 Phase, definition of mechanical, 14,  
     20  
 Phase angle, 328*ff*, 342  
 Phase transformation, 118  
 Photoelasticity, 154, 308  
 Photon, 44  
 Planck's constant, 33  
 Planck's function, 114  
 Planck's oscillator, 32, 114*ff*  
 Plastic deformation, propagation of,  
     350*ff*  
     theories of, 263*ff*, 272*ff*  
 Plastic flow, theories of, 263*ff*  
 Plastic hinges, 482*ff*, 490, 498*ff*  
 Plasticity, models of, 235, 247  
     of concrete, 524  
 Plastic potential, isotropic, 256, 265  
     of orthotropic body, 272  
 Plastic stability, in uniaxial compression, 500*ff*  
     of plates, 504*ff*  
 Plastic wave, propagation of, 350*ff*  
 Plastic zones, spreading of, 253*ff*,  
     425*ff*, 438  
 Plastometers, 409*ff*  
 Poisson's ratio, effect on yield condition of, 259  
     of work-hardened metal, 281, 303  
 Polymerization, 67, 81*ff*, 104, 400  
 Polymers, fracture of, 163  
 Polyphase materials, 105*ff*  
     damping of, 335  
     fracture of, 211*ff*  
     yield limit of, 197  
 Polystyrenes, 141  
 Potassium, 53*ff*  
 Potential, elastic, 215*ff*  
     plastic, 256, 265, 272  
 Potential trough, 74*ff*, 80, 84, 123*ff*  
 Power functions, use of, 206*ff*, 544  
 Precipitation, 101*ff*, 118*ff*, 128*ff*,  
     149*ff*, 286, 314  
 Prestressed concrete, 525*ff*  
 Primary bonds, 62*ff*, 74  
 Properties, types of, 17, 19, 94  
 Pseudosolids, 196  
 Psychophysical test, 412  
 Quantization of frequencies, 44*ff*  
 Quantum concept, 32*ff*  
 Quantum ladder of frequencies, 33  
 Quantum numbers, 37*ff*, 64  
 Quasidergodic, 111  
 Radiation, infrared, 32, 80  
 Randomization of structure, 142  
 Rate processes, 113, 140  
 Rebound (scleroscope) hardness, 539*ff*  
 Recovery, 140, 145*ff*, 297*ff*  
 Redundant metal structures, carrying capacity of, 481*ff*, 491*ff*, 497*ff*  
     shakedown of, 485  
 Relaxation, effect on fracture of, 372*ff*, 383, 395*ff*  
     of bolted assemblies, 322*ff*  
     relation of creep and, 196, 244, 322*ff*  
 Relaxation centers, 338, 341  
 Relaxation function, 243*ff*, 322*ff*  
 Relaxation spectrum, 338*ff*, 342*ff*, 375*ff*, 382  
 Relaxation test, 322  
 Relaxation times, 225, 245, 306*ff*, 338*ff*  
 Remembrance function, 341*ff*  
 Residual stresses, in plastic bending, 424*ff*  
     in thick-walled cylinder, 439  
     structural model of, 274*ff*  
 Resilience, 138*ff*, 143*ff*, 280*ff*  
 Resins, 411*ff*, 517*ff*  
 Resonance, 64*ff*, 68, 73  
 Resonance curve, 327  
 Retardation time, 224  
 Reversibility, 118, 194  
 Rotating cylinder (disk), limiting velocity of, 441*ff*  
 Rotations, components of, 217

Rubber, testing of, 411  
thermoelasticity of, 218

Safety, factor of, 480, 514*ff*

St. Venant element, 235

Scatter of fatigue data, 567*ff*

Scleroscope hardness test, 539*ff*

Secondary bonds, 74

Sedimentation, 399

Selenium, 65

Self-diffusion, 118, 140, 145*ff*

Sensitivity testing in fatigue, 571

Shale, mechanism of formation of, 396

Sharing of electrons, 64*ff*, 73

Shear strain, octahedral, 181*ff*, 258  
principal, 181  
pure, simple, 184

Shear stress, critical resolved, 129*ff*  
effect of hydrostatic pressure on  
critical, 156  
octahedral, 176, 258  
principal, 176

Shell cases, fragmentation of, 381*ff*

Silicates, 66

Silicon, 90

Silver chloride, 73

Skeleton structures, 19, 67, 105*ff*,  
217*ff*

Slip, 121*ff*  
blocking of, 120*ff*, 124, 130, 149*ff*,  
283*ff*, 298, 301  
direction of, 129  
distribution of, 130*ff*, 142*ff*, 249*ff*,  
282*ff*  
initiation of, 132, 144*ff*, 250*ff*,  
286  
laminar, 130*ff*, 140*ff*  
temperature effect on discontinuous,  
150*ff*, 298*ff*  
unfinished, 95, 124

Slip bands, 132, 282  
spacing of, 249, 282*ff*  
viscosity of, 127*ff*, 132, 249

Slip delay, 149*ff*, 428, 500  
effect on transition temperature  
of, 379*ff*  
in hexagonal crystals, 286

Slip planes, 128*ff*  
direction of, 129, 250, 453  
precipitation on operative, 120*ff*,  
124, 130, 149*ff*, 286, 298  
rotation of, 133*ff*

Slip systems, 129, 132*ff*, 140

Sodium, 40, 57, 69

Sodium chloride, 49, 73

Solid, definition of, 8, 196  
energy content of, 7, 82*ff*

Solid solution, 88, 92

Sols, 398

Specific density, 55

Specific heat, 82*ff*, 115

Spring constant, generalized, 330,  
342

State, mechanisms of change of,  
109*ff*, 116, 120*ff*, 152

Strain, definition of, 178, 511*ff*  
finite, 219*ff*  
logarithmic, 182*ff*

Strain aging, 120*ff*, 300*ff*

Strain anisotropy, 154

Strain deviator, 180  
invariants of, 181

Strain ellipsoid, 177*ff*, 187

Strain-energy function, 215*ff*

Strain hardening, mechanisms of,  
140*ff*, 287  
temperature sensitivity of, 145*ff*

Strain-rate-temperature parameter,  
137*ff*, 210, 321, 337

Strain softening, 141*ff*

Strain tensor, 179*ff*, 186  
invariants of, 180

Stress-concentration factor, 426

Stress concentrations, in concrete,  
plastic relief of, 524  
in fracture theories, 357, 364  
in metals, plastic relief of, 426,  
473, 549  
relief in creep design of, 509*ff*,  
515

Stress deviator, 174  
invariants of, 174

Stress distribution, in bending,  
plastic, 421, 464  
with creep, transient, 467*ff*

Stress function, 452*f*  
 Stress gradient, effect of, 15, 426,  
 428*f*  
 Stress prism, 429  
 Stress quadric, 173  
 Stress-strain diagram, adiabatic,  
 217, 349*f*, 354, 557  
 Stress tensor, 172*f*, 185  
 invariants of, 173*f*  
 Structural damage, 474*f*, 506*f*, 524*f*  
 Structural element, 6, 12  
 Structural models, 232, 235  
 Structure, cubic, 90*f*  
     diamond, 66  
     hexagonal, 65, 90  
     linear (threadlike), 65  
     plane (sheetlike), 65, 71*f*  
 Structure-sensitive properties, 19, 94  
 Substitutional alloys, 57*f*  
 Sulphur, 65, 93  
 Superlattice, 87*f*  
 Suspensions, 398*f*

Talcum, 71*f*  
 Tellurium, 65, 90  
 Temperature, Debye, 115*f*, 136  
     homologous, 83*f*  
     transition, 374*f*  
     velocity-modified, 137*f*, 201, 321,  
     337  
 Tensile strength, ultimate, 550  
 Tensor, overstress, 269*f*  
     rank of, 171, 179, 185  
     strain, 179, 186  
     stress, 172*f*, 185  
 Textile fibers, 71, 105  
 Textural stresses, 101, 144, 285, 297,  
 348  
 Texture, formation of, 140*f*, 287*f*  
 Thermal expansion, coefficient of,  
 75*f*  
 Thermal fatigue, 510  
 Thermal instability of structure,  
 145*f*, 284, 297, 313*f*  
 Thermal oscillations, amplitude of,  
 74  
 frequency of, 79, 116

Thermal softening, 140, 145*f*, 297*f*  
 Thermoelastic effects, in metals,  
 217  
     in rubber, 218*f*  
 Thick-walled cylinder, delayed  
 cracking of, 394*f*  
     limiting pressure in plastic, 431*f*,  
     438, 476  
     residual stresses in, 439  
 Thixotropy, 13, 142, 401*f*, 406, 412*f*  
 Tin, 70, 93  
 Transformation, austenitic, 57  
     martensitic, 93, 135  
 Transition temperature, 374*f*  
 Transition velocity, 376*f*, 557*f*  
 Trusses, carrying capacity of, 487*f*  
 Twinning, 134*f*  
 Two-phase systems, 106

Uncertainty principle, 44, 49  
 Undercooled liquid, 85, 104*f*, 196  
 Unit, of action, Planck's, 33  
     of flow, 83, 86  
 Unit cell, 89

Valence crystals, 64, 70  
 Valence electrons, 41*f*, 54, 63, 66*f*  
 Variational principles in plasticity,  
 267, 279*f*, 487*f*, 496  
 Velocity-modified temperature,  
 137*f*, 210, 321, 337  
 Velocity strain, 263  
 Vibration, modes of, 77  
 Viscoelastic analogies, 321*f*, 509,  
 516  
 Viscometers, 409*f*  
 Viscoplastic body, 221, 269, 403  
 Viscosity, 136*f*, 141, 196  
     measurement of, 2, 338, 409*f*  
     of grain boundaries, 101, 344*f*  
     of metals, 151*f*, 338  
     of slip bands, 127*f*, 132, 249  
     structural, 404  
     temperature dependence of, 137,  
     307  
 Viscosity coefficient, 86, 229  
     generalized, 330, 342

Viscosity coefficient, of asphalt, 410  
of concrete, 523  
of paints, 412

Volumetric deformation, effect of anisotropy on, 154*f*  
reversibility of, 153*f*

Wave, propagation of plastic, 351*f*

Wave equation, 45*f*

Wave function, 43*f*, 50, 67*f*

Wave particle, 43*f*

Weakest-link concept, in design, 475  
in fracture, 360*f*

Wedge, limiting pressure on plastic, 446*f*

Wire drawing, force of, 462, 471

Work hardening, mechanisms of, 140*f*, 287*f*

Work-hardening diagrams, discontinuity of, 150, 286, 298

Work-hardening relations, derivation of general, 290*f*

Working stresses, 477, 512*f*

Xenon, 41

Yield conditions, 255*f*  
effect of hydrostatic stress on, 262  
representation of, 258*f*

Yielding of support, effect on carrying capacity of continuous beam, 491, 517

Yield limit, 139, 197, 252, 300, 400, 405

Yield rectangle, Prager's, 277

Yield stress, in impact test, 556*f*  
in tension test, 548*f*

Zero elongation, cone of, 177

Zero intercept of creep curves, 307, 319









### **DATE OF ISSUE**

**This book must be returned within 3/7/14  
days of its issue. A fine of ONE ANNA per day  
will be charged if the book is overdue.**

---

--	--	--	--	--

Class No. 620.1128 Book No. 8921

Vol.

Author... Friedenthal.....

Title Inelastic Behavior of  
engg. materials & structures  
Acc No. 76309

---